DOCUMENT RESUME

ED 091 871 EA 006 257

TITLE	Evaluation: Elementary and Secondary Education. A Report to the Governor and Members of the Legislature Prepared Pursuant to Section 3602 of the Education Law.
INSTITUTION	New York State Education Dept., Albany.
PUB DATE	Mar 74
NOTE	213p.; A related document is EA 006 147
EDRS PRICE	MF-$0.75 HC-$10.20 PLUS POSTAGE
DESCRIPTORS	*Academic Achievement; Educational Assessment; Educational Programs; Elementary Education; *Evaluation Methods; Measurement Instruments; *Measurement Techniques; Program Evaluation; Secondary Education; State Action; State Surveys; Tests; *Test Selection; *Urban Education
IDENTIFIERS	*New York State

ABSTRACT

This report reviews the progress made by the New York State Education Department and local school districts in developing and implementing a comprehensive student evaluation program for elementary and secondary education. With the assistance of local school district personnel and technical consultants, the department concentrated its efforts on identifying the most pressing information needs and the characteristics of instruments adequately supportive of the decisionmaking tasks of today and tomorrow. The "interim" program first described in the report makes the best and the most systematic uses of existing methods and information to evaluate current categorical and general educational programs, and gives the best answers currently possible to pressing, performance-related questions being asked by the educational community. A "projected" program, described in the final part of the report, provides an examination of two state-of-the-art problems that must be overcome if the performance-related questions asked by educators, the State government, or the public at large are to be answered. (Page 140 may reproduce poorly.) (Author/JF)

ED 091871

EVALUATION ELEMENTARY AND SECONDARY EDUCATION

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Report to the Governor and Members of the Legislature Prepared Pursuant to Section 3602 of the Education Law

EA 006 257

The University of the State of New York
THE STATE EDUCATION DEPARTMENT
Albany, New York 12224
March 1974

THE UNIVERSITY OF THE STATE OF NEW YORK

Regents of the University (with years when terms expire)

1984	Joseph W. McGovern, A.B., J.D., L.H.D., LL.D., D.C.L., Chancellor	New York
1985	Everett J. Penny, B.C.S., D.C.S., Vice Chancellor	White Plains
1978	Alexander J. Allan, Jr., LL.D., Litt. D.	Troy
1972	Carl H. Pforsheimer, Jr., A.B., M.B.A., D.C.S., H.H.D.	Purchase
1975	Edward M. M. Warburg, B.S., L.H.D.	New York
1977	Joseph T. King, LL.B.	Shelter Island
1974	Joseph C. Indelicato, M.D.	Brooklyn
1976	Mrs. Helen B. Power, A.B., Litt.D., L.H.D., LL.D.	Rochester
1979	Francis W. McGinley, B.S., J.D., LL.D.	Glens Falls
1986	Kenneth B. Clark, A.B., M.S., Ph.D., LL.D., L.H.D., D.Sc.	Hastings on Hudson
1983	Harold E. Newcomb, B.A.	Owego
1981	Theodore M. Black, A.B., Litt. D.	Sands Point
1988	Willard A. Genrich, LL.B., L.H.D.	Buffalo
1982	Emlyn I. Griffith, A.B., J.D.	Rome

President of the University and Commissioner of Education
Ewald B. Nyquist

Executive Deputy Commissioner of Education
Gordon M. Ambach

THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
OFFICE OF THE PRESIDENT OF THE UNIVERSITY
AND COMMISSIONER OF EDUCATION
ALBANY, NEW YORK 12224

To the Governor and the Members of the Legislature
of New York State

At the 1973 Legislative Session you amended Section 3602 of the Education Law to include the following provision for elementary and secondary education evaluation:

" On or before February first, nineteen hundred seventy four, the commissioner shall report to the legislature on the manner in which the funds provided under this subdivision were spent, including but not limited to the names of the school districts which received funds and the amount, the disposition of funds to other agencies or by the department, the programs which were financed from the funds, and an evaluation of such programs and their relationship to programs funded with moneys provided by the federal government under the provisions of the Elementary and Secondary Education Act of nineteen hundred sixty-five and programs funded with moneys provided by special aid programs of the state of New York, and the extent to which such programs have been incorporated into the regular school curriculum and recommendations for future action. An amount equal to two per centum of the total appropriation for categorical urban aid shall be made available to the department of education from such total appropriation for the evaluation of such program and for the development and implementation of a comprehensive student evaluation program for elementary and secondary education."

I am pleased to submit a first report to you on the status and results of the evaluation to date.

I urge your review of the "interim" and "projected" programs described in this report. Your continued support of this statewide evaluation program is requested.

Respectfully submitted,

Commissioner of Education

TABLE OF CONTENTS

Chapter		Page
I	INTRODUCTION	1
	A. Questions Being Asked Of The Educational System	2
	B. An Overview Of The Report	3
	C. Summary Of Findings And Progress--The "Interim" and "Projected" Programs	4
II	DESCRIPTION OF PROGRESS TOWARD THE INTERIM PROGRAM	19
	A. Question #1: To What Extent Are New York State's Elementary And Secondary Students Successfully Mastering Basic Skills?	20
	1. Study of Local School Districts' Nationally Standardized Test Programs	21
	a. Purposes of the Study	21
	b. Procedures Used in the Study	23
	c. Results from the Portion of the Study Completed to Date	23
	1) Extent to which Schools Administer Nationally Standardized Tests	23
	2) Time of Year in which Schools Administer Nationally Standardized Tests	28
	3) Nationally Standardized Tests Commonly Used by Schools	30
	4) School District Expenditures for Nationally Standardized Achievement Tests	32

Chapter		Page
	5) School District Practices in Reporting Summarized Test Results	33
	d. Description of the PEP/Standardized Test Comparability Study	35
	e. Summary of the Study	36
2.	Comparison of Regents Examination Results in Biology, English, French, Geometry and Social Studies with Results from Commercially-Published Standardized Tests	37
	a. Purposes of the Study	37
	b. Procedures Used in the Study	37
	c. Results of the Study	40
	d. Cautions in Interpreting the Results	43
	e. Summary of the Study	45
3.	New York City Prototype Study of Urban Education and Elementary and Secondary Education (E.S.E.A.) - Title I Programs - School Year 1973-74	46
	a. Purposes of the Study	46
	b. Procedures Used in the Study	47
	c. Results from Data Presently Available	48
4.	Evaluation of Urban Education Projects - School Year 1972-73	60
	a. Purposes of Evaluation	60
	b. Review of the Need for Urban Education	60
	c. Procedures Used in the Urban Education Program - How Urban Education Projects are Initiated, Approved, Financed and Amended	62

Chapter		Page
	d. A Profile of Urban Education Projects	67
	e. Resource Distribution by Project Types	77
	f. Results - The Achievement of Students in 1972-73 Urban Education Programs	81
	5. Testing to Measure the Effectiveness of Bilingual Education	86
	a. Spanish-English Linguistics Skills Test Development	88
	b. State Bilingual Testing	89
	c. Testing of Science Skills when Instruction has Been in Spanish	90
B.	Question #2: What Types Of Instructional Programs (Or Projects In The Case Of Categorical Aid Programs) Are Most Effective For Students Of Different Characteristics?	91
	1. New York City Prototype Study of Urban Education and Elementary and Secondary Education (E.S.E.A.) - Title I Programs - School Year 1973-74	92
	a. Purposes of the Study	92
	b. Procedures Used in the Study	93
	c. Planned Types of Analyses - Spring and Summer 1974	97
	2. Evaluation of Urban Education Projects - School Year 1972-73	99
	a. Reading Laboratory Program	100
	b. Operation Reading Success	101
	c. Junior High Reading Laboratories	101
	d. Diagnostic Reading Program	101

Chapter		Page
	e. Cluster Team Program	103
	f. Home Study Center for Disadvantaged Children	103
	g. Prescribed Reading Essentials Program	104
	h. Uplifting of Basic Skills	105
	i. Secondary Education - Reading Programs	105
	j. Community School	106
	k. Bilingual Program	107
	l. Mathematics Laboratories	107
	m. Service to Early Elementary Children	108
	n. Bilingual Program	109
	o. Reading Diagnostic Center	109
	p. English as a Second Language in Nonpublic Schools	110
	q. Remedial Reading Clinic	110
	r. Corrective Mathematics	111
3.	Urban Education Performance Indicators Study - School Year 1973-74	112
	a. Purpose of the Study	112
	b. Procedures to be Used in the Study	112
	c. Anticipated Results of the Study	112
4.	Study of the Characteristics of High Performing Upstate Schools	113
	a. Purposes of the Study	113
	b. Procedures to be Used in the Study	113
	c. Anticipated Results of the Study	114

Chapter		Page
	5. Study of the Effect of Pupil Mobility Upon Pupil Evaluation Program (PEP) Results	115
	a. Purposes of the Study	115
	b. Procedures Used in the Study	115
	c. Results of the Study	116
	d. Summary of the Findings	120
III	DESCRIPTION OF PROGRESS IN DEVELOPING THE PROJECTED PROGRAM: A COMPREHENSIVE STUDENT EVALUATION PROGRAM	121
	A. State Of The Art Reviews	122
	1. Existing Testing Methodologies: Is the Educational System Succeeding?	122
	2. "State of the Art" Review - Availability of Disaggregated Information	128
	B. Question #1: To What Extent Are New York State's Elementary And Secondary Students Successfully Mastering Basic Skills?	130
	1. Desired Purposes of Tests	130
	2. Progress in Developing More Suitable Measurement Devices	134
	a. The Development of a New Test Methodology	135
	b. Other Uses for Data Which will be Furnished by the New Test Methodology	139
	c. Development of Basic Competency Examinations	144
	d. The Implementation of the Instructional Evaluation System	147
	e. Pilot Study of Instructional Criterion-Referenced Objective Tests	148

Chapter		Page
	C. Question #2: What Types Of Instructional Programs (Or Projects In The Case Of Categorical Aid Programs) Are Most Effective For Students of Different Characteristics:	152
	1. The Instructional Evaluation System - Overview of Purposes	154
	a. Instructional objectives	154
	b. Instructional Program Designs	156
	c. Utilizing Information for Decision-making (i.e. feedback)	157
	d. Computerized Support System	158
	e. Summary of the I.E.S. Components	159
	f. Evaluation of I.E.S.-Based Instructional Programs	162
	2. Progress in Implementing the Instructional Evaluation System	166
	D. Question #3: To What Extent Are The Best Instructional Planning, Management And Evaluation Practices Being Used At Different Levels Of The Educational System?	168
	1. Instructional Evaluation System	168
	2. Categorical Funding Study	169
	a. Purposes of the Study	169
	b. Procedures Used in the Study	173
	c. Preliminary Results of Study	181
	d. Limitations of these Data	189
	e. Summary of Findings	189
APPENDIX A		

List of Tables

Chapter	Number		Page
II	1	Amount of Standardized Reading Testing during 1972-73 in Grades 3-8 by Public Schools Returning Report Forms	24
	2	Amount of Standardized Mathematics Testing during 1972-73 in Grades 3-8 by Public Schools Returning Report Forms	24
	3	Number of Schools Administering Nationally Standardized Tests of Reading during 1972-73 and the Percentage Based on Number of Schools Returning Report Forms by Grade and Community Type	26
	4	Number of Schools Administering Nationally Standardized Tests of Mathematics during 1972-73 and the Percentage Based on Number of Schools Returning Report Forms by Grade and Community Type	27
	5	Percent of 3,177 Public Schools Reporting Reading Test Results by Month of Year in Which They Administered the Tests	29
	6	Percent of 3,037 Public Schools Reporting Mathematics Test Results by Month of Year in Which They Administered the Tests	29
	7	Percent of 3,177 Public Schools Reporting Reading Test Results by Name of Standardized Test Administered	30
	8	Percentage of 3,037 Public Schools Reporting Mathematics Test Results by name of Standardized Test Administered	31

Chapter Number		Page
9	Public School District Expenditures for Standardized Achievement Test Materials and Services in Grades K-12 during 1972-73	32
10	Public School District Practices in Reporting Summarized Results on State and Nationally Standardized Achievement Tests During 1972-73 by 720 (95%) of 758 School Districts Statewide	34
11	Methods Used to Report Summarized Results on State and National Standardized Tests to School District Residents by the 352 Districts Reporting Results on State Tests and the 152 Districts Reporting Results on National Tests during 1972-73 (Table 10)	34
12	Description of the Reference Groups	39
13	Percentile Ranks of Selected Regents Examination Scores in State and National Norms	40
14	Mean Regents Examination Scores of State and National Norms Samples and Percentile Rank in State Norms	41
15	Correlations Between Regents Examination and Standardized Test Scores	41
16	Regents Examination Means and Standard Deviations for Reference Groups and for 1972 Norms Samples	45
17	Distributions in Urban Education and Title I Programs of Most Qualified, Qualified, and Unqualified Students	58
18	Approved Funding by Source in Urban Education District 1972-73	65
19	Urban Education Project Modifications	66

Chapter	Number		Page
	20	Number of Individuals Who Participated in Planning at the District Level	67
	21	Project Time Span	68
	22	Average Duration of Urban Education Offerings	69
	23	Years of Project Operation	69
	24	Ethnic Group	70
	25	Participants in Five Major Components by Type of School	71
	26	A Comparison of Actual Participation by Project Component between the Urban Education Program and ESEA, Title I Program	72
	27	Number of Direct Staff Employed by Level and Type in Urban Education Project Activities	74
	28	Number of Urban Education Projects Employing Paraprofessionals	75
	29	Number of Types of Paraprofessional Services Used in Urban Education	76
	30	Component Activity Costs for the Component Activity Combination with the Five Highest Expenditures	77
	31	Comparison of Percentages of State Approved Budgets by Project Component for Categorically Aided Programs in Urban Education Districts 1972-73	78
	32	Percent of State Approved Budgets, by Budget Category, For Categorically Aided Programs in Urban Education Districts, 1972-73	80

Chapter	Number		Page
	33	Types of Evaluators	83
	34	Number and Percent of Project Component Subgroups Showing Improvement by Subject Matter of Instruction	85
	35	Number of Participants and Weighted Mean by Project Component Category	86
	36	Individual Skills in Reading Comprehension - Percent of Right Answers for Comprehensive Test of Basic Skills, Reading Comprehension (Level 1)	95
	37	Grade Equivalent, National Percentile, and District Percentile Based Upon Comprehensive Test of Basic Skills, Reading Comprehension (Grade 5.2)	96
	38	Districts, School Buildings, and Pupils in the Sample	116
	39	Relative Frequency of Third Grade Transfer Pupils	117
	40	Comparative Test Results for Transfer and Non-Transfer Students	118
	41	Correlation Between Percent of Target Group Pupils in Total Enrollment and Percent of Target Group Among Non-Transfer Pupils	119
III	42	Grade Level Sample for Participating Districts	149
	43	Mean Reading Achievement Before and After ISS in Guilderland Central Schools	163
	44	Mean Mathematics Achievement in Guilderland Central Schools	164
	45	Percent of Students Achieving Below Aptitude in Reading and Mathematics	165

Chapter	Number		Page
	46	Reliability Coefficients of Indicators	177
	47	Mean Application Quality Scores by District and Program Type	181
	48	Combined Title I and Urban Education Application Quality Scores Over Time	182
	49	Combined Title I and Urban Education Evaluation Quality Scores by District Type Over Time	187

List of Figures

Chapter	Number		Page
III	1	Schedule of Study Activities (1973-74)	150
	2	The Instructional Evaluation System (IES) As a Closed Loop Feedback System	161
	3	Application Quality by District Type and Time	183
	4	Application Quality by District Type and Program	184
	5	Evaluation Quality By LEA Type and Time	188

I. INTRODUCTION

New York State government officials from the executive and legislative branches have engaged in increasingly frank and open discussions of the problems associated with supporting local school district efforts to meet the educational needs of the citizenry. This constructive dialogue has sometimes been incorporated into printed reports, such as the study on "Urban Education Evaluation Reports for the Legislature" in 1972, and the "Preliminary Proposal for 1974 Legislation on State Aid for Elementary and Secondary Schools" this year. This report to the Governor and Legislature on the Evaluation of Elementary and Secondary Education is expected to foster further mutual cooperation among all parties interested in high quality education at the lowest reasonable cost. The Report sets forth the Department's conception of how best to proceed to insure this objective, along with the actions taken during the past year consistent with this conception, and it invites the people and their elected representatives to review the Department's direction and the support it merits.

This report reviews the progress the New York State Education Department and local school districts have made in developing and implementing a Comprehensive Student Evaluation Program for elementary and secondary education. On May 17, 1973 the Governor and Legislature amended New York State's Education Law to provide funds for the program developed by the Department (with advice from local districts) over the past three years. The report describes the work to date and indicates the plans to continue and expand these efforts during the fiscal year 1974-75.

In this introduction, three questions being raised both within and outside the educational community are identified. Because of two national "state of the art" problem areas which present barriers to answering these questions, the Department has developed both a long-term or "projected" program which addresses itself to overcoming the "state of the art" problems, and a near-term or "interim" program which provides the best possible answers given the existing constraints.

The remainder of the introduction provides an overview of the "interim" and "projected" programs and summarizes the findings and the progress that has been made to date within both programs. A complete discussion of the "interim" program is presented in Chapter II. A similar detailed discussion of the "projected" program is presented in Chapter III.

A. QUESTIONS BEING ASKED OF THE EDUCATIONAL SYSTEM

The intent of all the work activities presently in progress or being planned with the funds presently allocated is to answer three performance-related questions being raised both within and outside of the educational community. The questions are: [1/]

A. To what extent are New York State's elementary and secondary students successfully mastering basic skills?

1/ The questions listed are very similar to those in the Legislative Commission on Expenditure Review in their report--Program Audit 7. 1. 72, June 30, 1972--entitled Urban Education Evaluation Reports for the Legislature (the reader of this report might be interested in reviewing the Education Department's response, pages 19-29 of that document).

B. What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

C. To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?

It is generally believed that answering these three questions in a fair and meaningful way for all school districts in New York State is presently feasible. The Department, as a result of analyses over the past three years, is convinced that it is developing a program, which is described in Chapter III, that will adequately answer the questions. However, the Department is also convinced that an "interim" program must be used until the necessary developmental and pilot work is completed which overcomes the "state of the art" problems identified in Chapter III.

B. AN OVERVIEW OF THE REPORT

The Department has concentrated its efforts during the past year in identifying, with the assistance of local school district personnel and technical consultants, the most pressing information needs and the characteristics of instruments (tests, surveys and the like) adequately supportive of the decision making tasks of today and tomorrow. This report describes the achievements already realized, and the ways in which these accomplishments are intended to serve as building blocks for even more significant progress in the future.

The "interim" program described in Chapter II makes the best possible uses of existing methods and the most systematic use of existing information to evaluate current categorical and general educational programs, and gives

the best answers currently possible to the three pressing, performance-related questions being asked from within the educational system and from the Governor, the Legislature, and the public at large. Conclusions in Chapter II have to be viewed with caution and interpreted carefully because of the existing "state of the art" barriers discussed in the beginning of Chapter III.

The "projected" program--A Comprehensive Student Evaluation Program--is described in Chapter III. Chapter III provides an examination of two "state of the art" problems which must be overcome if the three questions described above are to be answered in the future. The description of the "projected" program indicates how, as the various components are incrementally implemented statewide, the existing "state of the art" barriers can be overcome by: (1) improving testing methodologies so children are tested for specified and more useful purposes; (2) providing better information to educational decision makers and the public; and (3) improving instructional planning and decision-making structures.

C. SUMMARY OF FINDINGS AND PROGRESS--THE "INTERIM" AND "PROJECTED" PROGRAMS

Since many readers will not have the time to review this entire report, a brief summary of the findings and progress made in developing and implementing both the interim and projected programs is presented in this introduction. This summary is organized in a manner which begins with either the "interim" or "projected" program and then reviews the various work activities under each of the three performance-related questions already outlined under "A" above.

1. Description of Progress Toward the Interim Program (Chapter II)

a. Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

1) Study of Local School Districts' National Standardized Test Programs: For many years many individuals both within and outside the Department have expressed a desire to know more about local school districts' evaluation programs especially regarding their use of national standardized tests as a means for answering question #1 above. Because of this interest, the Department, as part of the 1973-74 interim program, designed and is implementing a study: a) to obtain information on the scope (i.e. extent to which schools administer national standardized tests, which tests are used, when the tests are administered, and what use are results given in terms of reporting within the districts) of tests being used; b) to determine the extent of the total financial commitment to such programs; c) to explore the potential of using results on nationally standardized tests as a possible alternative to rather than an adjunct to the Pupil Evaluation Program (PEP), and d) to determine the achievement levels of pupils in New York State schools in comparison with national norms. In the data analyzed at this time, answers to "a" and "b" are presented in this report. Answers to the remaining two components of the study will be released in May 1974 when a substudy which correlates 1973 PEP scores with two national standardized test norms will be completed.

The following is a summary of the findings included in this report.

Reports were submitted by 94% of all public schools statewide enrolling 95% of all public schools statewide. During the 1972-73 school year, 90% of these schools administered a reading test in one or more grades, testing 75% of their total grade 3-8 enrollment; 86% administered a mathematics test in one or more grades, testing 61% of their enrollment. About 75% of the schools using standardized tests report administering them in the spring months as compared with 25% in the fall. Most of the schools use the Metropolitan Achievement Tests, the Iowa Tests of Basic Skills, or the Stanford

Achievement Tests. Even though most school districts did return questionnaires regarding the scope of their testing program, there are variations in: a) how much testing different types of districts administered; b) time of testing; and c) tests used.

School districts reported spending a total of roughly $2.4 million on standardized achievement testing in grades K-12 during 1972-73, an average of about $.70 per pupil enrolled. Most school districts report summaries of their State standardized test results to school board and school staff members and about half the districts report State standardized test (PEP) results to district residents. Less than half the districts report summaries of results on nationally standardized tests to their school board and school staff members and only 21% report nationally standardized test results to district residents. School districts that report test results to district residents use district newsletters, public meetings and formal releases to news media with about equal frequency.

2) Comparison of Regents Examination Results in Biology, English, French, Geometry and Social Studies with Results from Commercially-Published Standardized Tests: The administration of Regents examinations have a long history in New York State. Many educators and students have questioned whether they should be continued and others have suggested that commercially-published standardized tests replace Regents exams since they would allow the State to compare its students with national norms. The Board of Regents have concluded that the examinations should be continued until better examinations become available. The Regents believe that standards must be set at the State level. It should be understood that Regents examinations are generally only taken by students with average and above-average scholastic ability. This study examines the feasibility of using the commercially-published tests in place of Regents examinations and also attempts to answer the following two additional questions:

   1. How do the achievement test scores of New York State Regents-level students compare with scores obtained by high school students nationwide?

2. How closely correlated are Regents examination and standardized test scores in any given subject?

The results of the study indicate that the achievement of New York State students in June of 1972 was generally superior to that of students nationwide in English and French and approximately equal to the achievement of students nationwide in biology, geometry and social studies. The study also concluded that the correlations between Regents examinations and standardized tests scores range from .67 to .77. About half of the variance of one test is predictable from the variance of the other.

In examining the question as to whether it appears advisable to substitute these national standardized tests for Regents examinations, the report indicates that after analyzing the advantages and disadvantages of both tests that the Regents examinations will be retained until better tests are available at a reasonable cost.

3) New York City Prototype Study of Urban Education and Elementary and Secondary Education (E.S.E.A.) - Title I Programs - School Year 1973-74: As part of the total evaluation effort, the Department reviewed all the problems associated with relying exclusively on local evaluation reports and concluded that it should test the feasibility of developing and implementing a Department (as opposed to district) managed evaluation program. In the work completed to date, pretest scores from the California Test Bureau/McGraw-Hill Comprehensive Test of Basic Skills - Form Q, Level I (reading comprehension section) in reading for all grade five students in a sample of ten New York community districts have been analyzed to provide answers to the following questions. Both the questions and the findings are summarized here.

The first question is--Are the ten districts qualified to be Title I and Urban Education Districts? The data from the pretest indicates that the answer is "yes" in that in a normal population of students taking a standardized test one should find that 50 percent of the students are scoring at the 50th percentile or below. In the sample of all fifth graders (24,172), 17,403 students or 72 percent of the students were found to be below the 50th percentile.

The second question addressed is--Are all qualified students in either E.S.E.A. - Title I or Urban Education Programs? The data indicate that two-thirds of those likely to benefit (i.e. those below the 50th percentile) from both compensatory programs are not being served by either of these programs. The evidence is clear that the compensatory funds now available are not sufficient to serve all of the students in need if need is defined as all students below the 50th percentile.

The third question addressed is--Are unqualified students in either the Urban Education or E.S.E.A. - Title I Programs, and if so is it at the expense of qualified students? The answer to the first part of this question is "yes" if unqualified students are defined as those having pretest scores at or above the 50th percentile. There are 678 fifth-graders in this category. If these individuals were replaced by students below the 50th percentile, an additional 4 percent of the students in need could be served.

The fourth question analyzed is--Are some students more qualified than others for entry into E.S.E.A. - Title I programs by virtue of their educational deficiencies? In defining who might be considered more qualified, the Department divided the scores below the 50th percentile into two groups. Those scoring in the 1-25 percentile range were arbitrarily defined as "most qualified" and those in the 26-50 percentile range as "qualified". The pretest scores indicate that of the 72 percent of the students scoring below the 50th percentile, 42 percent scored in the bottom quarter (1-25 percentile range) and 30 percent in the second quarter (26-50 percentile range).

The last question addressed is--Is systematic preference for admittance to E.S.E.A. - Title I or Urban Education Programs given to students in the "qualified" group or the "most qualified" group, or do preference decisions appear to be random? The evidence in this report suggests that relatively few students (678 students) from the unqualified category (i.e. over the 50th percentile) are in the two compensatory programs. Clearly these figures indicate that there is no concerted effort to fill these programs with unqualified students. Nor do these numbers suggest that unqualified students

are even being selected randomly along with other more qualified students. It may be human error that some unqualified students are provided with compensatory services, or it may be that the costs of perfection in targeting compensatory funds are simply prohibitive. In terms of comparing the "qualified" and the "most qualified" data, there does not appear to be any clear-cut preference pattern between these two groups in either compensatory program. Arguments can be made that the "most qualified" group should receive all the services but on the other hand it could also be argued that greater improvements could be made with the 26-50 percentile group. These issues cannot be resolved until more data are available.

4) Evaluation of Urban Education Projects - School Year 1972-73: This section of the report reviews: the purposes of the Urban Education Program; 1) how projects are initiated, approved, financed, and amended; 2) a profile of Urban Education projects in terms of (a) the degree and form of community involvement in planning, (b) whether the projects were regular school year, summer only, or twelve month projects, (c) the average duration of offerings, (d) enrollment by ethnic group and project type, (e) type of participant (i.e. public versus nonpublic and expected versus participating), (f) staffing patterns in projects, and (g) component activity costs; 3) how resources are distributed between the Urban Education Program and other categorical aid programs and within programs.

The last portion of this section of the report provides an overview of the methodology used in evaluating and results of Urban Education projects. The results are clearly encouraging in that the sample of projects used for the analyses indicates that for all components that the actual average monthly gain from the projects was .94 as compared with a predicted gain of .61.

5) Testing to Measure the Effectiveness of Bilingual Education: The Regents Position Paper on Bilingual Education places a high priority on this area of need. Since more compensatory funds are now being targeted to Bilingual Education programs, the Department has committed itself to improving evaluative information for these programs. Specifically, efforts are now

underway: 1) to measure the degree to which linguistic skills taught to students, whose major or only language is Spanish, are being successful in equipping the student to read, write, understand, and speak in English; 2) to develop valid tests which measure student achievement in the areas of mathematics, social studies and science when instruction is given in Spanish; and 3) to evaluate a sample of Bilingual and English as a Second Language programs which have a promise of success.

b. Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

1) New York City Prototype Study of Urban Education and E.S.E.A. - Title I Programs - School Year 1973-74: As previously indicated this prototype study was undertaken to collect improved information for evaluating compensatory programs. In the remainder of this present school year the following activities will be completed. A post-test will be administered to all grade five students in the ten community districts in New York City which are included in the Department's sample. In addition, an "experimental version" of the reading effectiveness measure will be administered to all students. In addition, the results of these tests (including the pre-test scores) will be used along with detailed information collected on variables for each compensatory program to conduct the following analyses: 1) a calculation of the actual growth of students in each program and an evaluation of the results relative to the growth expected on the basis of actuarial data obtained from a national norming sample; 2) a calculation of the extent to which students are able to comprehend text equivalent in difficulty to that which they are using in instruction; and 3) data will be reviewed to determine the feasibility of analyses of the following types: (a) program versus no program designs; (b) within-district program comparisons; (c) using both Basic Education Data (BEDS) - derived indicators and data from the Categorical Funding Study described previously for making between-district comparisons. The availability of this wealth of information for the first time will make the selection of specific studies to be undertaken a difficult one.

2) Evaluation of Urban Education Projects - School Year 1972-73: Under question #1, this summary has already provided most of the information regarding the rationale for the procedures used and the results of Urban Education projects. To answer question #2--What types of projects are most effective for students of different characteristics? --pages 100-111 of Chapter II provide a summary of "exemplary" projects.

3) Urban Education Performance Indicators Study - School Year 1972-73: For several years, the Department has been developing the Performance Indicators in Education (P. I. E.) project as a statistical method of determining how various characteristics of the district relate to achievement. Generally, the data used in the statistical model has been BEDS (Basic Education Data System) information and PEP (Pupil Evaluation Program) scores. As part of the evaluation for the present year, this work is being extended to examine specific Urban Education projects. The purpose is to develop Urban Education performance indicators to determine how various characteristics of projects relate to achievement.

4) Study of the Characteristics of High Performing Upstate Schools: The purpose of this component which will be developing in the coming year, is to identify characteristics of upstate schools which are highly successful in promoting student achievement given the conditions under which they operate and to identify practices in schools which are correlated with higher than anticipated student achievement in the basic skills.

5) Study of the Effect of Pupil Mobility Upon Pupil Evaluation Program (PEP) Results: Each year, after the PEP test results have been reported to the schools, it is not uncommon to have principals complain that because they have a large pupil turnover, they are being unfairly "blamed" for the educational deficiencies of pupils who have been in their schools for only relatively short periods of time.

Of course, what the general public fails to realize is that it is not at all a matter of "blame". The primary purpose of the PEP program is to identify and locate a target group of pupils for special educational programs.

If pupils are properly classified in the target group, on the basis that their scores fall below the Statewide Reference Point established for that purpose, how long the pupils have been enrolled in the school is not a particularly important consideration for purposes of educational planning. The point remains that students in the target group have certain educational needs, and the number of target group pupils in a school is an index of unmet needs. The validity of this index is in no way diminished by the fact that more or less of the pupils are transferees.

Nevertheless, it is true that there has been a tendency in some sectors of the press and the public to equate the size of the target group in a particular school with the effectiveness of the educational program. Certainly that error would be compounded if pupil mobility were ignored. It was determined then that it would be useful to explore the dimensions of the mobility problem, and the relationship between pupil mobility and PEP scores.

The results of this study indicate the following. The frequency of transfer students, defined as students who began enrollment in Grade 1 in another school, is fairly high even by the time that the students reach Grade 3. In the State as a whole, it is estimated that about 29% of the Grade 3 students tested in the October 1972 PEP program were transfer students. As a group, the transfer students did in fact tend to have lower PEP reading test scores than the non-transfer students. However, this student mobility factor did not seem to have a sufficiently differential effect on the PEP test results of schools in the same community type to affect the relative standing of individual schools. That is, if transfer students had been eliminated from the PEP reports, the relative standings of individual schools in Grade 3 reading scores, as compared with other schools of the same community type, would have remained virtually unchanged.

2. Description of Progress in Developing the Projected Program: A Comprehensive Student Evaluation Program (Chapter III)

a. Review of two "State of the Art" Problems--(1) Existing Testing Methodologies; Is the Educational System Succeeding, and, (2) Availability of Disaggregated Information

The Department, after analyzing existing testing methodologies, has concluded that standardized tests are not as suitable as new forms of criterion-referenced tests for determining whether New York State's elementary and secondary students are successfully mastering basic skills. In addition, the Department is convinced that computer support systems are needed to be able to determine which instructional programs are more successful than others and to assist local school districts in improving instructional program planning, management and evaluation practices.

b. Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

1) Desired Purposes of Tests: The Department has concluded that three levels of tests are required to fulfill educational decision-making needs. First, there is a need to determine how well a student performs on tasks that may be encountered in adult life. Second, the test(s) must be able to measure progress toward adult competence over grades, and also be sensitive enough to measure individual and program progress within a period such as a school year. Third, there is a need for test(s) to verify that specific subskills result in a given level of competence at a specified point in time, and to verify the most effective sequences in which subskills can be acquired.

2) Progress in Developing More Suitable Measurement Devices: The emphasis of the work completed or in progress regarding the first two levels of testing needs described above has been in the priority area of reading. A one-year research effort has been completed. The results of this effort were: (1) a determination of the properties desired in the new testing methodology; (2) a description of the development tasks which must be undertaken to develop a new reading effectiveness measure; (3) a review of existing approaches to the

measurement of the effectiveness in reading, both in terms of the properties desired and the technical problems associated with meeting the specified properties; and (4) a proposed work plan for developing the actual tests. An experimental version of a new reading effectiveness measure will be administered to all grade five students in ten New York City community districts in the spring of 1974. To continue the development work on the reading effectiveness measure, the Department has submitted a two-year proposal to the National Institute of Education (N. I. E.). Other funding will be sought this spring if N. I. E. funding is not secured. If additional funds are available, work on a similar measure for mathematics will be undertaken. The Department believes that the development of these tests will be useful in identifying some of the factors contributing to less-than-desired levels of effectiveness. A review of analyses possibilities is included in Chapter III.

The Department is making significant progress in implementing criterion-referenced tests to satisfy the need for tests which measure subskills, the third level of testing identified. This is being accomplished by implementing the Instructional Evaluation System described in a subsequent section.

3) Plan to Satisfy the Immediate Need for Basic Competency Examinations: The reading effectiveness measure will not be operational for two years because of the necessary work required to sample adult and school materials, develop an improved readability criterion scale, and an improved predictive readability formula. Consequently, the Department has examined short term programs because it has committed itself in the near term to producing basic competency examinations in reading and mathematics.

One alternative would be to utilize a measure of minimum adult competence developed by the Educational Testing Service. A second alternative would be to utilize an experimental version of the reading effectiveness measure. With this alternative a sample of materials used in the school districts would be made and their readability calculated utilizing existing readability scales. The materials' readability would

be accepted as a de facto standard. A third alternative being considered is utilizing instructional objectives and matching criterion-referenced test items. A decision as to which alternative is to be used will be made shortly.

4) Pilot Study of Instructional Criterion-Referenced Objective Tests: Throughout this report, the Department indicates that standardized tests are not as useful as criterion-referenced tests at the instructional level. To determine the feasibility of supporting criterion-referenced testing on a broad scale in all the schools in New York State, a pilot study has been undertaken during this school year (1973-74) in ten school districts with different reading program needs in grades 4-6.

The study has already yielded some important data on the feasibility of using criterion-referenced testing on a statewide basis. It is now clear that the participating districts have largely selected the same instructional objectives in the areas of vocabulary and comprehension. Where the districts generally vary is in the "content" associated with the objectives and in the number of testing levels (e.g. levels 4-6) applicable to the children in a given classroom or building. The reading objective bank developed by the Department appears to be adequate for the task and many useful test items are being developed which can help other districts avoid doing this work into the future. In 1974-75, this study will be expanded in the same districts to cover grades 1-6 in reading.

c. Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

1) Instructional Evaluation System: The Department has made meaningful progress in developing and implementing the Instructional Evaluation System. This system consists of two major components which when implemented together in any local school district make it possible to determine what instructional programs are more successful than others.

The first component of the system is a set of procedures at the local level for: (1) systematically developing instructional program objectives that can be operationalized across classrooms; (2) developing or assembling instructional activities and criterion-referenced test items linked to instructional program objectives; (3) measuring performance with test items, and interpreting test scores using predetermined performance criteria; (4) systematically reallocating instructional resources (e.g. materials, teacher time) to improve student performance; and (5) systematically recording instructional activities, objectives, and test scores to provide a basis for upgrading instructional program content and evaluating instructional performance over time.

The second component is a computerized information-handling component which scores tests, keeps records on pupil and programs and generates reports for students, parents, teachers and educational decision-makers. Two computerized software systems are presently being used to implement the Instructional Evaluation System--the Instructional Support System (I.S.S.) which was developed and implemented in one New York State upstate district and the Comprehensive Achievement Monitoring (C.A.M.) system which was developed at the national level.

The Department has decided to use the I.S.S. in New York City and C.A.M. in upstate and Long Island districts. The Instructional Evaluation System must be implemented gradually because of the work involved for any given district. In the present year the highest priority is to begin working with New York City community districts. One district has already committed itself to implementing I.S.S. in all of its thirteen elementary schools for mathematics in grades 1-6. Other community districts are also considering utilizing the system. In upstate and Long Island districts, C.A.M. is already being used in many districts. Efforts are underway to expand the coverage through B.O.C.E.S. personnel.

This report includes some preliminary data indicating that the system is having an impact. (See pages 162-166).

d. Question #3: To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?

1) Instructional Evaluation System: The implementation of the Instructional Evaluation System has provided the Department with data concerning how instructional planning, management and evaluation practices can be improved. It is clear that the availability and use of the two components described above leads to an examination of how existing practices at the local school district level should and can be improved. Findings in this regard will be made available in subsequent reports.

2) Categorical Funding Study: The Department has preliminary findings from a special study which is designed to systematically examine instructional planning, management and evaluation practices in a sample of school districts presently receiving funds for compensatory education programs (i.e. E.S.E.A. - Title I and Urban Education). The Department believes that the recommendations in a final report in the fall of 1974 will be generally applicable to all educational practices, not just to those connected with compensatory efforts.

The goal of the study is to recommend feasible, cost-effective means for the State Education Department to improve the planning, management and evaluation of compensatory programs for the educationally disadvantaged. There are six components to the study which are reviewed in Chapter II. Preliminary findings of one component are presented in some detail. These findings have to do with the project proposals and evaluation reports submitted to the Department for approval. These findings will be analyzed further after field interviews are completed. The scaling of the project applications with measures (called application and evaluation quality measures) attempted to determine how district's planning, management and evaluation practices (as submitted on proposals and evaluation reports) scored in relationship to contemporary planning, management and evaluation standards.

The scores indicate that there has been an upward trend in application quality over the four-year period studied for all district types and for both the Urban Education and E.S.E.A. Title I projects. Similar improvement was found in evaluation scores for New York City community districts. Evaluation scores for large and moderate upstate districts remained relatively stable over the period of the study. Despite the existence of these improvements in scores, however, the scores were still found to be quite low when compared to contemporary planning, and evaluation standards. This indicates that the final recommendations of the study will have to deal with this finding if through field interviews these initial findings are verified.

II. DESCRIPTION OF PROGRESS TOWARD THE INTERIM PROGRAM

In the introduction to this report, the summary of the two "state of the art" problem areas--existing testing methodologies and the availability of disaggregated information--indicated that the Department is constrained in its ability in the short run to answer completely the three performance-related questions being asked of the educational system by the Governor, the Legislature, the public at large and educators themselves for current categorical and general educational programs. The Department's efforts described in this Chapter, however, even though not as suitable as those described in Chapter III-the projected program, have dramatically improved over those reported in the past. The Department is confident that it has implemented an interim program which makes the best possible use of existing methods and the most systematic use of existing information to answer the three questions addressed in the introduction.

In this chapter work efforts will be reviewed under the two performance-related questions used in the previous chapter. The third question--"To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?"--is discussed in Chapter III.

Under question #1--"To what extent are New York State's elementary and secondary students successfully mastering basic skills?"--the following work efforts are reviewed in this section: 1) Study of Local School Districts' Nationally Standardized Test Programs; 2) Comparison of

Regents Examination Results in Biology, English, French, Geometry and Social Studies with Results from Commercially-Published Standardized Tests; 3) New York City Prototype Study of Urban Education and Elementary and Secondary Education (E.S.E.A.) - Title I Programs - School Year 1973-74; 4) Evaluation of Urban Education Projects - School Year 1972-73; and 5) Testing to Measure the Effectiveness of Bilingual Education.

Under question #2--"What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?"--the following work efforts are reviewed: 1) New York City Prototype Study of Urban Education and E.S.E.A. - Title I Programs - School Year 1973-74; 2) Evaluation of Urban Education Projects - School Year 1972-73 (Exemplary Projects); 3) Urban Education Performance Indicators Study - School Year 1972-73; 4) Characteristics of High Performing Upstate Schools; and 5) Study of the Effect of Pupil Mobility Upon Pupil Evaluation Program (PEP) Results.

The following sections review each of these efforts under each of the two questions.

A. QUESTION #1: TO WHAT EXTENT ARE NEW YORK STATE'S ELEMENTARY AND SECONDARY STUDENTS SUCCESSFULLY MASTERING BASIC SKILLS?

In reviewing the efforts described below it is important to keep in mind the reasons why each of these work efforts has been included in this chapter describing the interim program. There are, as indicated in the introduction, barriers imposed by testing methodologies and the unavailability of disaggregated

information. These barriers will not be reviewed again here but they should be considered in reading what follows. It should also be noted that some of the statistical procedures used here and under question #2 are presently being reviewed in terms of their appropriateness.

1. Study of Local School Districts' Nationally Standardized Test Programs

a. Purposes of the Study

New York State, with its Pupil Evaluation Program (PEP), is one of the few states that has mandated administration of uniform State-developed tests. The PEP tests are administered annually in reading and mathematics in all schools at the beginning of grades 3, 6 and 9. Results from these tests are widely reported. In addition, many local school districts administer nationally standardized tests. These tests are used at the local level for various purposes including that of determining how well their students are mastering basic skills.

For years many individuals both within and outside the Department have expressed a desire to know what specific tests are being used in local districts, at what time during the school year they are administered, how many districts administer such tests, what the cost is to the schools, etc. The Department, therefore, designed and implemented a study to address these questions. The specific purpose was to obtain information on the scope of the evaluation activities now going on in the local school districts, and the extent of the total financial commitment to such evaluation programs. While there has been increased public demand in recent years for more evaluation and accountability at the State level, it was hypothesized in undertaking this study that there has been inadequate recognition of the extent to which local school districts have been actively and continuously concerned with evaluation.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

A second purpose of the study was to explore the potential of using local school results on nationally standardized tests as an alternative to the PEP testing program. The major purposes of PEP are to help identify and locate the low-achieving students in the State and to determine trends in the general level of achievement from year to year. While the PEP tests have heretofore not permitted comparison of the level of achievement in New York State with national norms, comparability studies for that purpose are now under way, so that a new and additional subpurpose of the PEP tests may be considered to be a comparison of New York State test scores with national norms. The key question, of course, is whether local school tests would be equally as effective as the PEP results in serving the particular purposes for which the PEP tests were designed.

It may be that, despite variations in tests used by local schools, the grade level scores of the students so obtained will identify low-achieving students as effectively as the PEP. Also, despite variations in practice that may exist with respect to the particular tests, or the time of year at which the tests were administered, it is possible that the results may be sufficiently comparable so that if they are integrated on a systematic basis. they will provide a useful picture of the trend in Statewide achievement levels from year to year, or in comparison with national norms.

If that were found to be the case, local tests would have certain advantages over PEP. Local tests could provide data for all grades rather than merely three grades, they would be more comprehensive in scope and probably better suited to evaluating school programs, they would permit greater flexibility with respect to date of testing, and duplicate testing would be reduced.

Accordingly, the survey of local school testing included an analysis of the test results as well as the testing practices. To accomplish this second purpose of the survey, however, further analysis will be necessary in order to determine

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

comparability with results provided by PEP. In particular, when the PEP comparability study (see page 35) is completed in May, it will be possible to verify whether a Statewide grade-level obtained by pooling a variety of different local school tests administered at different times to varying percentages of students would be comparable to the Statewide grade-level obtained from the PEP tests.

b. Procedures Used in the Study

During the first week of September 1973 official report forms were sent to all public and nonpublic schools in the State requesting that they report the names of standardized tests of reading and mathematics administered gradewide during the school year 1972-73 in any of grades 3 to 8, inclusive, along with the numbers of pupils tested, the dates administered, and the school medians for total scores. At the same time, public school districts were requested to respond to a brief questionnaire covering standardized achievement test expenditures and reporting practices for grades K-12 in the school district as a whole.

The requested return date for the completed forms was October 12, 1973. Schools and school districts were directed to submit reports even if no tests had been administered. Where returns were not received by the due date, follow-up was made by telephone or by letter. However, a cut-off date of November 30 had to be established for purposes of computer processing.

c. Results from the Portion of the Study Completed to Date

1) Extent to which Schools Administer Nationally Standardized Tests: Tables 1 and 2 show that 3,177 public schools, 90% of those returning report forms, administered grade wide one or more standardized tests of reading to pupils in their grade 3-8 enrollments, and 3,037 (86%) administered one or more in mathematics. These schools tested roughly 75% of their combined grade 3-8 enrollment with a reading test and 61% with a mathematics test.

TABLE 1: Amount of Standardized Reading Testing during 1972-73 in Grades 3-8 by Public Schools Returning Report Forms

	Schools Returning Forms				Schools Reporting Results			
Grades	No.	% of Total	Gr. 3-8 Enroll.	% of Total	No.	%	No. Tested	%
3	2700	95%	259,789	96%	2050	76%	193,191	74%
4	2637	95	261,810	96	2300	87	218,596	83
5	2563	95	259,757	96	2180	85	211,989	82
6	2226	95	257,926	96	1867	84	205,220	80
7	967	91	250,777	94	636	66	161,429	64
8	946	91	240,995	94	560	59	150,803	63
Comb.	3535	94%	1,531,054	95%	3177	90%	1,141,229	75%

TABLE 2: Amount of Standardized Mathematics Testing during 1972-73 in Grades 3-8 by Public Schools Returning Report Forms

	Schools Returning Forms				Schools Reporting Results			
Grades	No.	% of Total	Gr. 3-8 Enroll.	% of Total	No.	%	No. Tested	%
3	2700	95%	259,789	96%	1916	71%	178,152	69%
4	2637	95	261,810	96	1599	61	132,918	51
5	2563	95	259,757	96	2074	81	200,521	77
6	2226	95	257,926	96	1817	82	200,264	78
7	967	91	250,777	94	381	39	80,961	32
8	946	91	240,995	94	500	53	137,313	57
Comb.	3535	94%	1,531,504	95%	3037	86%	930,129	61%

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

Of the schools administering tests, the highest percentage administer reading tests in grades 4, 5 and 6 (87%, 85%, 84%); and mathematics tests, in grades 5 and 6 (81% and 82%). Fewer schools administer tests in grades 7 and 8. In these grades, reading tests are administered by 66% and 59%, and mathematics tests by only 39% and 53%. The percentages of pupils tested in the individual grades in these schools are generally lower but have a pattern roughly comparable to the percentages of schools using tests.

Tables 3 and 4 provide an analysis of the amount of testing in public schools by type of community. The combined community type totals in Table 3 show that schools in most every type, except Large City (77%), administer a standardized test in reading to pupils in one or more grades of their grade 3-8 enrollments. The New York City schools administer a reading test in each grade in practically every school, but in schools in other community types practices vary. For example, schools in medium size cities administer reading tests in most every grade except grade 7; the percentages for these medium size city schools range from 85% to 99% in grades 3-6 and grade 8, but only 52% administer a reading test in grade 7. In large cities, 81% of the schools administer a reading test in grade 7, but only 33% in grade 8. In village-suburban and rural areas, schools tend to test reading in each of grades 3-6, but only half of the schools in these types test reading in grades 7 and 8 (Vill-Sub., 59% and 56%; Rural, 50% and 46%).

In mathematics (Table 4) as in reading, the combined totals for community type show that most schools, except those in large cities, administer a mathematics test in one or more of the grades in their grade 3-8 enrollments. The grade-by-grade patterns within community types are similar to those for reading, but generally show slightly less gradewide testing in mathematics than in reading. For example, New York City schools administer mathematics tests in grades 3, 5, 6 and 8 only; no gradewide tests are administered in grades 4 and 7.

TABLE 3: Number of Schools Administering Nationally Standardized Tests of Reading during 1972-73 and the Percentage Based on Number of Schools Returning Report Forms by Grade and Community Type

Grade	NYC	Lg.City	Med.City	Sm.City	Vill-Sub	Rural	Combined
Third Grade							
No.Schools	615	103	81	160	705	386	2,050
% of Tot.	99%	45%	85%	64%	67%	85%	76%
Fourth Grade							
No.Schools	613	158	95	206	835	393	2,300
% of Tot.	99%	70%	99%	86%	83%	88%	87%
Fifth Grade							
No.Schools	609	136	94	190	773	378	2,180
% of Tot.	99%	62%	98%	83%	80%	87%	85%
Sixth Grade							
No.Schools	502	102	92	161	644	366	1,867
% of Tot.	98%	50%	96%	87%	80%	87%	84%
Seventh Grade							
No.Schools	171	70	13	40	185	157	636
% of Tot.	98%	81%	52%	71%	59%	50%	66%
Eighth Grade							
No.Schools	158	28	24	34	174	142	560
% of Tot.	98%	33%	96%	61%	56%	46%	59%
Combined							
No.Schools	783	210	110	274	1251	549	3,177
% of Tot.	99%	77%	99%	89%	89%	85%	90%

TABLE 4: Number of Schools Administering Nationally Standardized Tests of Mathematics during 1972-73 and the Percentage Based on Number of Schools Returning Report Forms by Grade and Community Type

Grade	NYC	Lg.City	Med.City	Sm.City	Vill-Sub	Rural	Combined
Third Grade							
No.Schools	617	99	66	144	631	359	1,916
% of Tot.	99%	43%	69%	57%	60%	79%	71%
Fourth Grade							
No.Schools	0	154	90	194	780	381	1,599
% of Tot.	0%	68%	94%	81%	77%	85%	61%
Fifth Grade							
No.Schools	610	136	79	175	704	370	2,074
% of Tot.	99%	62%	82%	76%	73%	85%	81%
Sixth Grade							
No. Schools	503	103	87	153	616	355	1,817
% of Tot.	98%	50%	91%	83%	76%	84%	82%
Seventh Grade							
No.Schools	0	62	10	30	152	127	381
% of Tot.	0%	72%	40%	54%	49%	41%	39%
Eighth Grade							
No.Schools	160	23	19	29	152	117	500
% of Tot.	99%	27%	76%	52%	49%	38%	53%
Combined							
No.Schools	784	209	103	256	1168	517	3,037
% of Tot.	99%	77%	93%	83%	83%	80%	86%

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

2) Time of Year in which Schools Administer Nationally Standardized Tests: Tables 5 and 6 provide information on the months during which public schools administer nationally standardized tests of reading and mathematics. Most of the 3,177 schools reporting reading reading test results for one or more of grades 3-8 administer nationally standardized tests of reading in the spring, 34% in April, 37% in May and 7% in June --a total of 78%. About 28% of the schools administer tests in the fall, 9% in September, 16% in October and 3% in November. The combined percentages, of course, add to more than 100% because some schools administer two or more reading tests per year on a gradewide basis and a school is included in each of the months in which it administered a test. These 3,177 schools were 90% of the 3,535 returning report forms and 85% of the 3,751 schools with enrollments in these grades.

Of the 3,037 schools reporting mathematics test results, most schools generally administer mathematics tests at times during the year similar to those for reading, but with one notable exception. Although, as for reading, a high percentage of schools administer mathematics tests in the spring, only 9% administer them in April and 60% in May as compared with 34% and 37% administering reading tests in these months. These 3,037 schools were 86% of the 3,535 schools returning report forms and 81% of the 3,751 schools with enrollments in these grades.

TABLE 5: Percent of 3,177 Public Schools Reporting Reading Test Results by Month of Year in Which They Administered the Tests

Month (No:=)	Gr.3 (2050)	Gr.4 (2300)	Gr.5 (2180)	Gr.6 (1867)	Gr.7 (636)	Gr.8 (560)	Combined (3177)
September	8%	8%	7%	7%	7%	6%	9%
October	9	13	13	12	19	18	16
November	2	3	2	3	3	4	3
December	0	0	1	0	2	2	1
January	2	2	2	4	4	5	4
February	2	2	2	4	2	2	3
March	3	4	4	3	2	2	5
April	36	35	36	36	32	34	34
May	39	36	34	33	27	26	37
June	7	5	6	5	7	6	7
Total(1)	108%	108%	107%	107%	105%	105%	119%

1 Possible for total percentages to exceed 100% because schools administering tests more than once a year, e.g., fall and spring, were counted each time tests were administered.

TABLE 6: Percent of 3,037 Public Schools Reporting Mathematics Test Results by Month of Year in Which They Administered the Tests

Month (No.=)	Gr.3 (1916)	Gr.4 (1599)	Gr.5 (2074)	Gr.6 (1817)	Gr.7 (381)	Gr.8 (500)	Combined (3037)
September	7%	10%	7%	6%	5%	4%	8%
October	10	18	14	12	24	18	16
November	1	4	2	3	5	4	3
December	0	1	1	0	3	2	1
January	2	3	2	4	5	5	4
February	1	2	1	3	1	2	3
March	3	5	3	4	3	2	5
April	7	11	8	9	7	5	9
May	69	47	61	58	41	55	60
June	6	7	5	5	8	4	5
Total(1)	106%	108%	104%	104%	102%	101%	114%

1 Possible for total percentages to exceed 100% because schools administering tests more than once a year, e.g., fall and spring, were counted each time tests were administered.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

3) Nationally Standardized Tests Commonly Used by Schools: Tables 7 and 8, giving the percentages of schools using various nationally standardized tests of reading and mathematics, show that the most widely used tests in the State are the Metropolitan Achievement Tests (40% and 41%), the Iowa Tests of Basic Skills (31% and 33%) and the Stanford Achievement Tests (19% and 19%). This is true for each of grades 3-8 and for all grades combined. All other tests combined have less than a 20% usage rate.

TABLE 7: Percent of 3,177 Public Schools Reporting Reading Test Results by Name of Standardized Test Administered

Name of Test (1) (No. =)	Gr. 3 (2050)	Gr. 4 (2300)	Gr. 5 (2180)	Gr. 6 (1867)	Gr. 7 (636)	Gr. 8 (560)	Combined (3177)
MAT	44%	41%	40%	39%	41%	39%	40%
ITBS	23	31	29	31	24	29	31
SAT	19	17	18	18	14	12	19
CAT	5	5	5	6	6	5	5
Gates-M	5	4	4	4	6	6	6
CTBS	3	3	3	2	1	1	2
SRA	2	1	2	2	3	3	2
STEP	0	0	0	0	1	2	1
Other	2	1	2	2	6	5	3
Total (2)	103%	103%	103%	104%	102%	102%	109%

1 MAT/Metropolitan Achievement Tests; ITBS/Iowa Test of Basic Skills; SAT/Stanford Achievement Test; CAT/California Achievement Test; Gates-M/Gates-MacGinitie; CTBS/Comprehensive Test of Basic Skills; SRA/SRA Achievement Series; STEP/Sequential Tests of Education Progress

2 Possible for total percentage to exceed 100% because schools administering two or more different tests in the same grade during the year, e.g. MAT in the fall and ITBS in the spring, were counted each time a different test was administered.

TABLE 8: Percentages of 3,037 Public Schools Reporting Mathematics Test Results by Name of Standardized Test Administered

Name of Test (1) (No. =)	Gr. 3 (1916)	Gr. 4 (1599)	Gr. 5 (2074)	Gr. 6 (1817)	Gr. 7 (381)	Gr. 8 (500)	Combined (3037)
MAT	46%	20%	41%	39%	20%	42%	41%
ITBS	25	44	31	32	42	32	33
SAT	20	23	18	19	22	15	19
CAT	5	7	5	5	9	5	5
CTBS	3	4	3	3	2	1	2
SRA	2	2	2	2	3	3	2
STEP	0	1	0	0	1	2	1
Other	1	2	1	1	1	1	1
Total (2)	102%	103%	101%	101%	100%	101%	104%

1 MAT/Metropolitan Achievement Tests; ITBS/Iowa Test of Basic Skills; SAT/Stanford Achievement Test; CAT/California Achievement Test; CTBS/Comprehensive Test of Basic Skills; SRA/SRA Achievement Series;STEP/Sequential Tests of Educational Progress

2 Possible for total percentage to exceed 100% because schools administering two or more different tests in the same grade during the year, e.g. MAT in the fall and ITBS in the spring, were counted each time a different test was administered.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

4) School District Expenditures for Nationally Standardized Achievement Tests: Expenditures for standardized achievement test materials and services during 1972-73 for grades K-12 by public school districts in the State were a little over 2.4 million dollars (Table 9). This estimate is based on reports received from 95% of all the public school districts in the State and assumes districts from which no reports were received had expenditures similar to those reporting within the same community type. The average expenditure statewide amounted to $.70 per pupil enrolled. Among community types, the per pupil enrolled expenditures ranged from lows of $.61 and $.62 in rural and large city districts to highs of $.73 and $.74 in New York City and Village-Suburban districts.

TABLE 9: Public School District Expenditures for Standardized Achievement Test Materials and Services in Grades K-12 during 1972-73

Community Type	No. of Districts	K-12 Enrollment	Total Expenditures[1]	Ave. Expend. Per Pupil Enrolled
New York City	31	1,122,788[2]	$ 814,173[2]	$.73
Large City	7	204,394	126,843	.62
Medium City	8	85,099	59,647	.70
Small City	46	218,988	140,695	.64
Vill-Suburban	263	1,398,014	1,032,419	.74
Rural	403	443,935	270,707	.61
Statewide	758	3,473,218	$2,444,484	$.70

(1) Assumes that school districts not reporting (5% statewide, Table 1) had per pupil expenditure rates comparable to those in other school districts in their community type.

(2) Includes total Central Board enrollments and expenditures as well as those for the 31 local districts

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

5) School District Practices in Reporting Summarized Test Results: It appears to be fairly common practice among school districts in the State to provide their school boards, their administrators and their teachers with summarized information on the results of New York State standardized tests (PEP). As shown in Table 10 the percentages are 81%, 89% and 83% respectively. Although roughly 90% of the school districts report using nationally standardized tests, only 40%, 55% and 49% provide their boards and administrative and teaching staffs with summarized results on these tests.

It should be noted that even though summaries are not provided to the teachers, they do receive results for each child in their classroom. These results do indicate that a greater effort is required on the part of the school districts who do not presently report on their results in making information available. However, if the school districts conclude that the results from such tests are not useful for decision-making, a decision should be made to invest the monies now being spent in a testing program which does give them a capability to make improved decisions.

About half (49%) of the districts provide their residents with summarized reports on State tests, and only 21% provide their residents with summarized information on national tests. These figures again indicate that school districts should make a greater effort in reporting to community. The State results appear to be more useful from their point of view. If this is the case they should make greater efforts to interpret these results for the community.

Table 11 shows that the school districts that do report test results to district residents seem to utilize a variety of methods to accomplish this purpose rather than rely on any one particular method, and some apparently use a combination of methods. About 50% report using each of the three primary methods available to them: District Newsletter, public meetings and formal district releases to the news media.

TABLE 10: Public School District Practices in Reporting Summarized Results on State and Nationally Standardized Achievement Tests During 1972-73 by 720 (95%) of 758 School Districts Statewide

Summarized Results Reported to	N.Y. State Tests No. of Districts	%	National Tests No. of Districts	%
Board of Education	586	81%	289	40%
School Administrators	640	89	398	55
Teachers	596	83	356	49
District Residents	352	49	152	21

TABLE 11: Methods Used to Report Summarized Results on State and National Standardized Tests to School District Residents by the 352 Districts Reporting Results on State Tests and the 152 Districts Reporting Results on National Tests during 1972-73 (Table 10)

Reporting Methods	N.Y. State Tests No. of Districts	% (N=352)	National Tests No. of Districts	% (N=152)
District Newsletter	168	47%	66	43%
Public Meetings	192	54	86	57
Formal District Release to News Media	183	51	76	50

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

d. Description of the PEP/Standardized Test Comparability Study

The PEP/Standardized Test Comparability Study has two major purposes: 1) to compare the achievement of pupils in New York State with the achievement of pupils nationwide in reading and mathematics at the beginning of grades 3 and 6; 2) to determine the correlations between the New York State PEP tests and commercially-published standardized tests in reading and mathematics for grades 3 and 6.

As a by-product of the first of these purposes, comparability tables will be developed which will enable schools to interpret PEP test scores in terms of national norms. Heretofore, only New York State norms have been provided for the PEP tests.

In order to accomplish these purposes, a sample of schools has been asked to submit individual pupil scores in reading and mathematics obtained on the PEP tests for grades 3 and 6 and on one of the two most commonly used commercially-published standardized tests, the Iowa Tests of Basic Skills and the Metropolitan Achievement Tests. The tests are those administered in the fall of 1973 as a part of the regular school testing program; schools were not asked to do any additional testing for this study. When the scores have all been received, the sample will be as follows:

Commercial Tests	No. of Pairs of Scores Grade 3	Grade 6
Iowa Tests of Basic Skills	4,000	5,500
Metropolitan	3,000	2,500

These scores will be obtained from a representative sample of approximately 30 school districts.

Comparability will be established by the equipercentile method. Scores having the same percentile rank in the equating sample will be considered comparable, permitting the interpretation of PEP test scores on the basis of the commercial test norms. Thus, it will be possible to compare the percent of pupils in New York State who fall below any given PEP test score with the percentage of pupils in the national norms sample who would fall below this score if the PEP test were administered nationwide.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

The correlation between the PEP tests and the commercial tests will indicate the extent to which the tests measure common factors. If the correlation is high and the tests are similar in content--.90 or higher--the test results could be considered interchangeable, and the same general interpretations about pupil achievement can be made from the PEP test results as from the results of the Iowa or the Metropolitan tests.

e. Summary of the study

Reports were submitted by 94% of all public schools statewide enrolling 95% of all public school pupils statewide. During the 1972-73 school year, 90% of these schools administered a reading test in one or more grades, testing 75% of their total grade 3-8 enrollment; 86% administered a mathematics test in one or more grades, testing 61% of their enrollment. About 75% of the schools using standardized tests report administering them in the spring months as compared with 25% in the fall. Most of the schools use the Metropolitan Achievement Tests, the Iowa Tests of Basic Skills, or the Stanford Achievement Tests. Even though most school districts did return questionnaires regarding the scope of their testing program, there are variations in: a) how much testing different types of districts administered; b) time of testing; and c) tests used.

School districts reported spending a total of roughly $2.4 million on standardized achievement testing in grades K-11 during 1972-73, an average of about $.70 per pupil enrolled. Most school districts report summaries of their State standardized test results to school board and school staff members and about half the districts report State standardized test results to district residents. Less than half the districts report nationally summarized test results to their school board and school staff members and only 21% report nationally standardized test results to district residents. School districts that report test results to district residents use district newsletters, public meetings and formal releases to news media with about equal frequency.

Decisions regarding whether or not to substitute commercial standardized tests for the PE Program will be made after a substudy is completed in the spring which will determine the correlation between PEP and two national standardized tests and the similarity of content between the tests.

2. Comparison of Regents Examination Results in Biology, English, French, Geometry and Social Studies with Results from Commercially-Published Standardized Tests

a. Purposes of the Study

The administration of Regents examinations has a long history in New York State. Many educators and students have questioned whether they should be continued and others have suggested that commercially-published standardized tests replace Regents exams since they would allow the State to compare its students with national norms. The Board of Regents have concluded that the examinations should be continued until better examinations become available. The Regents believe that standards must be set at the State level even though Regents examinations are generally only taken by students with average and above-average scholastic ability. This study examines the feasibility of using the commercially-published tests in place of Regents examinations and also attempts to answer the following two additional questions:

1) How do the achievement test scores of New York State Regents-level students compare with scores obtained by high school students nationwide?

2) How closely correlated are Regents examination and standardized test scores in any given subject?

b. Procedures Used in the Study

The procedures used to answer the first question involved the following steps:

1) Choosing five Regents examination subjects for study, and selecting commercially-published standardized tests in each of these five subjects.

2) Administering each standardized test to a sample of students (reference group) preparing to take the corresponding Regents examination in June 1973.

3) Establishing the comparability of selected scores in each pair of tests, Regents and standardized.

4) Finding the percentile ranks of these selected scores in two norm tables for each subject: (a) the norms for the June 1972 Regents examination, and (b) the norms for the standardized test.

The five Regents examination subjects selected for this study were biology, English, French, tenth year mathematics (geometry), and social studies. The standardized tests selected were:

1) Nelson Biology Test, Revised Edition, Form E, published by Harcourt Brace Jovanovich, 1965.

2) Stanford Achievement Test, High School Battery, Test 1: English, Form X, published by Harcourt Brace Jovanovich, 1965.

3) Pimsleur French Proficiency Tests, Test 3: Reading Comprehension, Second Level, Form C, published by Harcourt Brace Jovanovich, 1967.

4) Modern Geometry Test, published by Houghton Mifflin, 1971.

5) Stanford Achievement Test, High School Battery, Test 6: Social Studies, Form X, published by Harcourt Brace Jovanovich, 1965.

An effort was made to select tests which were reasonably similar in content to the corresponding Regents examinations, were recently standardized, and could be administered in one class period. In the case of English, French, and social studies, it was impossible to satisfy the first two of these criteria, although the tests finally selected came closer than any of the other tests which were available.

The standardized tests were administered by the regular classroom teachers in June 1973 prior to Regents examination week. Except for French, the tests were scored commercially, and the scores were reported to the Bureau of Elementary and Secondary Educational Testing. The French test was hand-scored in the Bureau. After the Regents examinations had been administered and scored locally, these scores were also forwarded to the Bureau where they were paired with the standardized test scores. Table 12 indicates the number of pairs of scores which were used in this study and the schools in which the students were tested.

Table 12

Description of the Reference Groups

School	Number of Pairs of Scores				
	Biology	English	French	Geometry	Soc. Stud.
Bethlehem, Delmar			85		
Binghamton Central				160	
Cazenovia	108	94		82	121
Dobbs Ferry	76				46
Far Rockaway, N.Y.C.			124		
John E. Glenn, Huntington	223			172	
Monroe Woodbury	176	166		203	188
Oceanside			128		
Total	583	260	337	617	355

The equipercentile method was used to determine comparability: scores having the same percentile rank were considered comparable. Three points in each distribution of Regents scores and one point in each distribution of standardized test scores were selected for study. They were (1) the Regents examination score of 90, which is considered a high score in each Regents examination; (2) the mean Regents examination score obtained in June 1972; (3) the Regents examination score of 65, which is the minimum passing score; and (4) the mean standardized test score obtained at the time of standardization.

To answer the second question, the pairs of scores used to establish comparability were also used to compute Pearson product moment correlations.

c. Results of the Study

Table 13 shows, for each subject, the percentile ranks of selected Regents examination scores in the two sets of norms. The national percentile ranks are approximations based upon the use of comparable scores. It can be seen that for all but three pairs of scores the percentage of students with lower scores is greater for the national norms sample than it is for New York State students. This indicates that the achievement of New York State students in June 1972 was generally superior to that of students nationwide. The percentile ranks of the remaining three scores suggests that the achievement of New York State students at these levels in social studies and geometry was approximately equal to the achievement of students nationwide.

Table 13

Percentile Ranks of Selected Regents Examination Scores in State and National Norms

Regents Score	Percentile Rank									
	Biology		English		French		Geometry		Soc. Stud.	
	St.	Nat.	St.	Nat.	St.	Nat.	St.	Nat.	St.	Nat.
90	80	91	90	96	90	96	80	80	85	84
Mean	45	55	50	74	45	67	45	56	50	50
65	24	31	14	36	17	40	28	40	13	18

A similar picture is presented in Table 14, which approximates the Regents examination score of the "average student" nationwide in each subject and compares this score with the mean score attained by the sample of students used to compute the June 1972 percentile equivalents. In each subject the mean score attained by students in the New York State norms sample exceeds the estimated mean score of students in the national norms sample. Ignoring percentile rank differences of less than fifteen points, one would conclude that students in New York State achieved at a higher level than students nationwide in English and French, and at the same level in biology, geometry, and social studies.

Table 14

Mean Regents Examination Scores of State and National Norms Samples and Percentile Rank in State Norms

Subject	State Norms Sample		National Norms Sample	
	Mean	%ile Rank	Mean	%ile Rank
Biology	75	45	74	45
English	75	50	66	20
French	74	45	68	30
Geometry	72	45	70	40
Social Studies	76	50	74	45

The correlations between Regents examination and standardized test scores are given in Table 15. This table indicates that there is a substantial correlation between scores obtained for each pair of tests, ranging from .67 to .77. About half of the variance of one test is predictable from the variance of the other.

Table 15

Correlations Between Regents Examination and Standardized Test Scores

Subject	r
Biology	.77
English	.71
French	.74
Geometry	.72
Social Studies	.67

In addressing the question as to whether the Regents examinations should be replaced by commercially-published standardized tests an analysis was undertaken to determine the advantages and disadvantages of each test. The following factors were considered:

1) Regents examinations are more closely related to the properties described in Chapter III--the "projected program--than standardized tests.

2) Regents examinations are new each time they are administered, which varies from one to four times a year; standardized tests are kept on the market for 5-10 years without revision.

3) Regents exams are entirely secure; standardized tests are readily available to classroom teachers.

4) Regents exams are based upon State-recommended courses of study; standardized tests are based on an analysis of textbooks which generally are composites of several courses of study.

5) Regents exams are 3-hours in length; standardized tests require about 40 minutes. (In French, for example, the Regents examination measures reading, writing, and listening skills, but the standardized test that the Department used measures only reading comprehension.)

6) The Regents examination costs are much lower than those from publishers. The Regents examination costs are based upon an average test development expenditure of $8,000 per examination and printing costs of $30 per thousand. The Regents examinations in four of the five subjects measured in the study (all except French) are among those used most widely and therefore those with the lowest per pupil costs. June Regents examinations are written by from about 2,500 pupils in such subjects as Italian and Hebrew to 35,000 in Earth Science and in Physics to over 140,000 in English, in Social Studies and in Biology. The average overall per pupil cost for June Regents examinations is in the $.25 to $.35 range and would be slightly higher for a total January, June and August examination program package.

The per pupil cost figures listed below are for testing the 2,000 pupils for this study.

Subject	Per Pupil Cost	
	Regents Exam	Standardized Test
English	$.10	$.73
Social Studies	.10	.73
Biology	.10	.51
Geometry	.11	.46
French	.32	.32

These figures illustrate how large some of the differences can be, but are not entirely typical because Regents examination costs are directly related to the number of tests written for each examination because the larger the number of tests written the lower is the per pupil examination cost.

7) The Department has not surveyed the usage of commercial tests in the field but based on its field experience is convinced that few school districts presently administer such tests. If a decision were made to replace the Regents examination, either the local school district or the State would have to allocate additional funds.

It is the Department's belief that the advantages of continuing the Regents examinations clearly outweigh the disadvantages until some other suitable substitute is available at a reasonable cost.

d. Cautions in Interpreting the Results

Studies seeking to evaluate the effectiveness of New York State secondary schools usually include an attempt to compare the achievement of students in New York State with the achievement of students in other states. Valid comparisons of this type are extremely difficult to obtain because of the number of variables that must be controlled. If, for example, a state that is especially successful in keeping its students in high school until graduation is compared, on the basis of achievement test scores, with a state having a high percentage of "dropouts", the state with many dropouts will be given a decided advantage because the students remaining generally are more successful. A similar advantage will accrue to a state that has, for one reason

or another, a large proportion of very able or highly motivated students. The problems encountered in interstate comparisons are analogous to those involved in comparing individual schools solely on the basis of achievement test scores. Nevertheless, if the variables which affect achievement test scores are assumed to be approximately equal for all states, it is possible to make some generalizations concerning pupil achievement in New York State vis-a-vis pupil achievement in the rest of the nation. Indeed, such generalizations are not uncommon.

The major finding of this study appears to be consistent with the findings of previous studies and with the evidence derived from national testing programs: New York State students achieve at a level which is at least equal to or, in some subjects, is superior to that of students nationwide. However, the absence of normative data for the June 1973 Regents examinations dictates caution in the acceptance of this conclusion. Although the technical characteristics of Regents examinations remain fairly stable from year to year, there are occasional exceptions. When the 1973 norms become available, this study can be revised to permit the placing of considerably greater confidence in the results.

It should also be noted that the comparability of scores established in this study is somewhat dependent upon the reference groups used, although comparability will hold reasonably well for other groups drawn from the same population. Table 16 shows the extent to which the reference groups are representative of the 1972 norms samples.

The substantial correlations found between scores on Regents examinations and scores on standardized tests do not, of course, suggest that the tests are interchangeable. Otis IQ scores correlate .71 with Nelson Biology Test scores (Nelson, 1965, p. 13), .77 with Stanford English scores, and .76 with Stanford Social Studies scores (Gardner, 1965, p. 21), but the use of the Otis IQ test in place of any or all of these achievement tests would not be considered. The first criterion in selecting a test must be content validity, and no high school achievement test can be used in New York State unless it is based upon the State-recommended course of study.

Table 16

Regents Examination Means and Standard Deviations
for Reference Groups and for 1972 Norms Samples

Subject	Reference Group			1972 Norms Sample		
	n	Mean	S. D.	n	Mean	S. D.
Biology	583	79.7	12.3	20,428	74.6	15.5
English	260	72.9	11.7	21,992	74.5	11.5
French	337	76.8	12.7	5,348	74.5	12.9
Geometry	617	75.5	17.9	20,351	71.6	19.3
Social Studies	355	78.6	10.8	18,874	75.8	12.7

e. Summary of the Study

In June 1973, samples of New York State high school students preparing to take Regents examinations in biology, English, French, geometry, and social studies were administered commercially-published standardized tests in these subjects. The results were used to determine the comparability of scores on each pair of tests, permitting the achievement of pupils in New York State to be compared with the achievement of pupils nationwide. Correlations between Regents examination and standardized test scores were also computed.

The results indicate that New York State students achieve at a level which is at least equal to or, in some subjects, is superior to that of students nationwide. Substantial correlations were found between the scores obtained on Regents examinations and standardized tests, ranging from .67 to .77.

The Department has concluded that the Regents examinations should be retained until better instruments are available at a reasonable cost.

3. New York City Prototype Study of Urban Education and Elementary and Secondary Education (E.S.E.A.) - Title I Programs - School Year 1973-74

a. Purposes of the Study

In the process of determining what work activities the Department should undertake in the school year 1973-74 a review of past practices was undertaken. It was determined that four general problem areas beyond those "state of the art" issues discussed in Chapter III of this report could and should be dealt with, realizing, however, that the limitation of funds available would prohibit the Department from undertaking any effort in all school districts receiving compensatory education funds.

The first problem identified was that tests given to measure the effect of programs are not comparable across districts or across projects either in terms of the test given or in the time when they are administered to students. Second, because of high levels of student mobility and the lack of computer ability at the local level it is generally not feasible to systematically trace pre and post test scores of individual students by project. Third, it was determined that the effects of compensatory projects are difficult to separate from the effects of the regular educational program available to all students because of the absence of comparable control groups. Lastly, it was determined that more detailed information was needed regarding the projects, if more detailed analysis were to be performed.

After considering these problems the Department concluded that a prototype study should be undertaken in New York City, since a majority of the compensatory funds are targeted there. This study was designed to test the feasibility of developing and implementing a Department (as opposed to district) managed evaluation program. The evaluation was also designed to further the development of the "projected program" described in Chapter III and to overcome the problems identified above.

Specifically, the purposes of this year's work activities in this prototype activity as they relate to Question #1 are to answer the following specific questions (the purposes related to question #2 are reported on pages 92-99).

1) Are the 10 community districts qualified to be E.S.E.A. - Title I and Urban Education Districts if a standardized test is used as an indicator of educational need?

2) Regarding the distribution of qualified and unqualified students in the 10 community districts:

   . Are all qualified students in either an E.S.E.A. - Title I or an Urban Education project?

   . Are unqualified students in these programs, and, if so, is it at the expense of qualified students?

3) Regarding differences among qualified students:

   . Are some qualified students more in need of compensatory aid than others?

   . Is systematic preference in targeting E.S.E.A. - Title I or Urban Education programs given to the "most qualified" group or to the "qualified" group (difference between "most qualified" and "qualified" follows), or are selection decisions apparently random?

b. Procedures used in the Study

The data used below results from the administration of a standardized test of reading (the California Test Bureau/ McGraw-Hill Comprehensive Test of Basic Skills - Form Q, Level I) to all grade five students in 10 New York City community districts. Grade five was selected since the Department's Reading Bureau concluded that results of any grade selected below that level might be overly affected by the reading approach used (i.e. if skills are taught in different sequences in various approaches, the test selected might or might not match that sequencing). The

CTBS test was selected because it provides historical data on what students nationwide have done on a pre and post test basis. Operationally, this means that after the post test is administered in the spring of 1974 it will be possible to determine whether students in compensatory projects have made greater gains in reading achievement than would be predicted from national performance data. (This will be reported in the fall of 1974.) A computerized student file has been established to insure that the post-tests are properly matched to the pre-tests. In addition, detailed information on the project has been collected. The types of program data and their use will be reviewed under Question #2.

An additional element of great interest to the participating districts will be included in a fall report. As indicated in Chapter II, the Department is in the process of developing a "reading effectiveness measure". In the fall of 1973, data were collected on what reading and social studies instructional materials are used in each classroom. The readability of these materials is presently being analyzed using existing readability formulas. In the spring of 1974 an "experimental version" of the reading effectiveness measure will be administered. In this experimental version, test item difficulty will be controlled and a quasi-cloze technique used for deleting words in passages which are ranked according to difficulty. As a result of this activity, it will be possible to determine whether or not students have met the reading "standards" set by their reading and social study materials, and to determine whether or not the grade-level difficulties of the materials correspond to the grade level reading ability of the students.

c. Results from Data Presently Available

For the first time, the data available for analysis in this ten district sample provide an unduplicated count of fifth graders, categorized on the basis of numbers enrolled in Title I programs, Urban Education programs, both programs or neither program. It also provides a distribution of these students' pretest scores on the reading comprehension section of the California Test Bureau/McGraw-Hill Comprehensive Test of Basic Skills - Form Q, Level I. The distribution, or percentile ranking, of scores was related to the national norms for this test and was used to determine how students taking this test do relative to one another, and to the original group of students upon whom the norms are based.

While interpretation of the enrollment by program data is fairly straightforward, interpretation of the test score data is considerably more difficult. Methodological difficulties present in standardized tests cloud any interpretation of test scores or percentile rankings--beyond comparison of students to each other or to the norm group. Yet other questions which one would hope testing instruments would address are clearly more fundamentally important.

For example, suppose the average score for a school district on this test is the same as the average score obtained by all students in the norming group, thereby placing the school at the 50th percentile. Can we say that, on the average, this school district has fulfilled its responsibility to teach students all that they should know at this point in their education? Clearly we cannot. All we know is that these students, on the average, did as well as the norming group. What we do not know is whether what the norming group did on the average was enough to warrant determination that their education was in some fundamental sense adequate. The answer to this question could only be provided by a criterion-referenced test. which had specified levels of competence by grade level, and which was systematically related to some well-defined "anchor" statement of skills necessary for the effective functioning of graduates in society.

Unfortunately, the state of the art in testing is not sufficiently advanced at this point to make usage of such tests practical. Hence, this analysis is forced to make the best interpretation possible from existing standardized test results and to accept the ambiguities in interpretation as are unavoidable.

In this analysis, a major area of ambiguity which cannot be avoided involves the determination of what constitutes a "qualified" as opposed to "unqualified" student for Title I and for Urban Education classification purposes. This is a critical determination because much of the potential usefulness of this type of data lies in the direction of being able to determine accurately the numbers of students being served by Title I and Urban Education programs relative to the total need for these programs.

Since no criterion-based instrument is available to determine how many students are failing to achieve at logical levels of competence for their age/grade level, we have to fall back on definitions of "qualified" and "unqualified" which are based on each student's standing relative to his peers and/or to the norm group. This ambiguity enters when students are designated as qualified solely because they achieve at levels below other students. It may be that in fact both groups of students (or neither group) are achieving adequately in terms of a steady progression toward reasonable adult competencies. This we do not know and, as a result, educators are forced to make some arbitrary classification based on relative percentile ratings of various groups of students.

Convention dictates that any student achieving below "grade level" (i.e. below the 50th percentile) be viewed as falling behind and warranting special attention. The Department will adopt that arbitrary convention and augment it somewhat by dividing this group into two parts: the first ("most qualified") and second ("qualified") quarters. The first quarter contains students in the 1-25 percentile rank range, and the second quarter those in the 26-50 percentile rank range. All students achieving above the 50th percentile will be taken to be "unqualified" or inappropriate targets of Title I - Urban Education programs.

Another and similarly arbitrary definition of "qualified" students includes those students achieving more than one year behind "grade level", i.e., below approximately the 33rd percentile on the California Test Bureau/McGraw-Hill test. This definition was rejected mainly because the available data are not categorized on this basis, and exposition would necessarily be more involved to accommodate this problem without any real additional benefits to be gained.

The following analysis will center on addressing several questions of an evaluative nature. The answers to these questions should provide interesting insights into the role of compensatory programs in the 10 districts in this sample. The questions to be addressed are as follows:

1) Are the 10 districts qualified to be Title I and Urban Education districts?

2) The distribution of qualified and unqualified students in the 10 district sample:

   a) Are all qualified students in either a Title I or Urban Education program?

   b) Are unqualified students in these programs, and if so, is it at the expense of qualified students?

3) Differences among qualified students:

   a) Are some students more qualified than others to be the targets of compensatory programs?

   b) Is systematic preference in targeting Title I or Urban Education programs given to the most qualified group or the qualified group, or are selection decisions apparently random?

These questions will be addressed in the analysis which follows.

(1) Are the Ten Districts Qualified to be Title I and Urban Education Districts?

The data indicate that 72 percent, or 17,403 of the 24,172 fifth graders in the ten district sample, were scoring at the 50th percentile* or below on the California Test Bureau/McGraw-Hill Comprehensive Test of Basic Skills. This indicates that 22% more of the fifth graders in these districts were achieving below grade level (i.e., below the 50th percentile), than would be expected if achievement

* Achievement at the 50th percentile means a grade equivalent score of 5.0 on this test. However, this test was given in November almost three months after the start of the school year. In this sense, grade level should be about 5.3, or above the 50th percentile. Hence the inclusion of 50th percentile students in the qualified group.

levels in these districts matched the national achievement norms specified for this test. On the assumption that the minimally desired achievement level for these students is the average achievement level for the norming group, clearly the ten districts sampled qualify in the aggregate as having exceptionally large numbers of underachievers.

On a district-by-district basis, however, the variation in the proportion of students achieving at or below the 50th percentile ranges from a low of 53.6 percent to a high of 84.8 percent, with all but one district having more than 60 percent of its students in this category.

There appear to be two major factors which can account for a district close to national norms in achievement becoming a recipient of Urban Education and Title I funds. First, eligibility is determined in part on the basis of individual school building needs, not just on the school district's overall situation. Hence, a district which, overall, may be achieving at national norms might still be eligible if one school in the district was in clear need of compensatory resources. Secondly, compensatory aid traditionally has been allocated on the basis of both underachievement and poverty, rather than just underachievement alone.

(2) The Distribution of Qualified and Unqualified Students in the Ten District Sample

(a) Are all qualified students in either E.S.E.A. Title I or Urban Education programs?

As was noted above, 72 percent or 17,403 of the 24,172 fifth graders in the ten district sample appear to be qualified to benefit from

Title I and Urban Education programs. Enrollment in Title I and Urban Education programs is 6,393 or only 37 percent of the total qualified student population in the ten districts. However, this overstates the extent to which the need in these districts are being addressed since some of these students in these programs appear unqualified for them on the basis of achievement. Subtracting out the students in these programs who are achieving at levels above the 50th percentile leaves only 5,715 qualified students being served. This figure represents only 33 percent of the total number of qualified students. Overall, then, a full two-thirds of those students likely to benefit from Urban Education and/or Title I programs are not being served by either of these programs.

In terms of individual districts in the ten district sample, there is a variation of between 11.3 percent and 65.3 percent in the qualified students being served by these programs, depending on which district is being reviewed. This suggests that between 34.7 percent and 88.7 percent of the total need for the Urban Education and E.S.E.A. - Title I programs at the fifth grade level is currently unmet, depending on the district being considered.

(b) Are unqualified students in either the Urban Education or E.S.E.A. Title I programs, and if so, is it at the expense of qualified students?

If unqualified students are defined as those pretesting at or above the 50th percentile, the answer is: yes, there are unqualified students in the Urban Education and Title I programs in the ten district sample. 678 fifth-graders fall into this category. If they were replaced by qualified students, the total number of qualified students being served would increase to 6,393 or 37 percent of the total pool of qualified students. This would mean that an additional 4 percent of the qualified students would be served by these programs.

This figure can be viewed from another, more positive perspective: that district efforts to insure that available places in Urban Education and Title I programs are filled by students that the programs were intended to serve have been 96 percent successful. However, even if the districts had been 100 percent successful in this regard, the special needs of a full 63 percent, or almost two-thirds of all fifth graders achieving below grade norms would remain unmet by Urban Education and Title I programs in the ten districts.

The number of unqualified students in individual districts of the ten district sample varies considerably. Viewed as a proportion of the total number of qualified students in each district, the number of unqualified students in Title I and Urban Education programs

varies from 0.5 percent to 9.9 percent of this total, depending on which district is being examined. This, of course, suggests that an additional 0.5-9.9 percent of the students qualified for these programs in individual districts could be served by them if these districts were totally successful in targeting only on the appropriate students for these programs.

(3) Differences Among Qualified Students

(a) Are some students more qualified than others for entry into Urban Education and E.S.E.A. Title I programs by virtue of their educational deficiencies?

Of the 72 percent of the students in the ten district sample who scored at or below the 50th percentile on the pretest, and hence warranted a designation as likely to benefit from Urban Education and/or Title I programs, 42 percent scored in the bottom quarter (1-25 percentile range). The remaining 30 percent scored in the second quarter (26-50 percentile range). Clearly those 42 percent achieving in the bottom quarter have a more clearly demonstrated need for compensatory program assistance, and from this perspective are more qualified for compensatory programs than the 30 percent of students in the second quarter. For purposes of further analysis we will designate the students in the bottom quarter as "most qualified" and the students in the second quarter as simply "qualified".

(b) Is systematic preference for admittance to E.S.E.A. Title I or Urban Education programs given to students in the "qualified" group or the "most qualified" group, or do preference decisions appear to be random?

Establishing priorities among different groups that stand to benefit from exposure to compensatory education programs is a complicated problem. Should the limited resources available be concentrated on those most qualified for assistance because they have the greatest need? Or should they be concentrated among students in the 26-50 percentile range on the assumption that this group, because their problems are more limited, are more in a position to benefit from compensatory programs? Whichever way this decision is made, it would clearly be unreasonable to orient compensatory programs toward students who are achieving above norms for their age/grade group.

Table 17 provides an overall picture of how students were selected for Title I and Urban Education programs in the ten district sample. In Table 25, data are presented to show the proportions of "most qualified", "qualified" and "unqualified" students in the total 5th grade enrollment, the Urban Education program enrollment, the Title I program enrollment, and the enrollment in both Urban Education and Title I programs.

The evidence suggests that relatively few students from the unqualified category are in the two compensatory programs. They number 678 or about 10.6 percent of the total number of students in these compensatory programs. Clearly these figures indicate that there is no concerted effort to fill these programs with unqualified students. Nor do these numbers suggest that unqualified students are even being selected randomly along with other more qualified students.

Comparison among the most qualified group and the qualified group, however, does not in general yield such clear-cut results, except for the small group of students in both Urban Education and Title I programs. Almost 90% of these students are from the most qualified group, indicating a clear preference pattern in terms of student selection for this joint program effort. The results for students in either Urban Education or Title I programs are not as clear. With 70.0 percent of all Urban Education students and 64.3 percent of all Title I students coming from the most qualified strata, clearly the general emphasis tends to be on this type of student. However,

Table 17

Distributions in Urban Education and Title I Programs of Most Qualified, Qualified, and Unqualified Students

Program	Most Qualified (1-25 Percentile)		Qualified (26-50 Percentile)		Unqualified (51-99+ Percentile)		Total 1-99+ Percentile)	
	Percent	Number	Percent	Number	Percent	Number	Percent	Number
Urban Education	70.0	1,387	21.7	430	8.3	164	100.0	1,981
Title I	64.3	2,743	23.8	1,015	11.9	508	100.0	4,266
Both Urban Education and Title I	88.4	129	7.5	11	4.1	6	100.0	146
Number not being served by compensatory education programs		5,869		5,820		6,090		17,779
Total Enrollment	41.9	10,128	30.1	7,276	28.0	6,768	100.0	24,172

the fact that a full 21.7 percent of the Urban Education and 23.8 percent of the Title I students are from the (less educationally disadvantaged) qualified strata indicates there probably is no clear-cut preference pattern between these groups used in targeting Urban Education and Title I programs.

Ranking these programs, the process for selecting students to be in both Urban Education and Title I programs placed heaviest emphasis (88.4 percent) on selection of students from the most qualified group. Urban Education was next with 70 percent of the enrollment being selected from students in the most qualified category. Title I fell last with only 64.3 percent of the program enrollment falling in the most qualified category. Title I also had the largest percentages of both qualified (23.8 percent) and unqualified (11.9 percent) students in the program.

If the targeting process was totally oriented toward selection from only the most qualified group of students, 2,134 more places in Urban Education and Title I programs would have been available for students in the most qualified group. This would have meant that 36.4% of the most qualified students now in neither of these programs could have been served in one of them.

In summary, the analysis presented above and that to follow in the spring are significant improvements over what has been done in the past. In addition, this prototype will allow similar improvements under Question #2 reported on pages

4. Evaluation of Urban Education Projects - School Year 1972-73

a. Purposes of Evaluation

Since the Governor and Legislature have a special interest in the State's Urban Education Program, this section of the overall report includes more descriptive information than other sections. One of the purposes, then, of this section is to review: why the Urban Education Program was established, why the project approach was adopted, and how projects are initiated, approved, financed and amended. In addition, an overall profile of projects and a report on how the resources are distributed by program type are presented. Lastly, the success of urban education projects is reviewed. This last component addresses itself to question #1--to what extent are New York State's elementary and secondary students successfully mastering basic skills? Specific exemplary projects are included under question #2 on pages

b. Review of the Need for Urban Education

Educational problems that result strictly from poor education usually can be cured with improved education. However, educational problems that result from related circumstances such as poverty, poor housing, employment, or pervasive anxiety about matters involving money or lack of it may not always be solved with improvements in education. Thus those problems which stem from society's general need for food, clothing or health may not be completely solved through education in general or the Urban Education Program specifically.

The major intent of the Urban Education legislation was to alleviate educational problems associated with poverty. From the view of the educator, one problem stood above all others: in 1967 when the Urban Education Program was funded, students, and especially those with a background of

poverty, were falling further behind grade level with the passing of each year. It became obvious that efforts were required to reverse this trend if meaningful progress toward equalizing educational opportunity was to be made. The Urban Education Program was an attempt to meet this need.

Estimates of the number of students participating in Urban Education programs provide some indication of the need for compensatory programs such as Urban Education. During the 1972-73 academic year there were over 500,000 duplicated participants in the Urban Education program. The number of duplicated participants is detailed elsewhere in the report. However, it should be noted that this figure represents but a small portion of the total number of students with poverty-related education deficiencies. It is presented mainly to show that the magnitude of the need for compensatory education is substantial. It should be noted that participants in Urban Education projects range in age from the prekindergarten child to the unemployed or underemployed adult who has not received a high school diploma or passed the high school equivalency test.

The Department has never, in the past, had firm figures from each district relative to the unduplicated count of participants in the Urban Education program or other categorically aided programs. In that regard, a form for eliciting such data from school districts for use with the 1973-74 report has been devised. Not only will such a report deal with the unduplicated count of participants across projects within a given program (e. g. the State Urban Education Program) but also across programs (E. S. E. A. - Title I and Urban Education) within a given district. The indicators from the preliminary 1973-74 data which were reviewed in the previous section strongly support the contention that Urban Education programs are serving only a small portion of those students who could stand to benefit from them.

A number of alternative strategies for addressing the problem of Urban Education have been considered over the years. New forms of school financing were recommended as a means of solving the problems as were adjustments in local control, integration, busing, and compensatory education. While the preceding alternatives all have potential for addressing certain urban problems, the designers of the Urban Education Program felt that compensatory education was especially promising as a strategy for equalizing educational opportunity.

Stated succinctly, if the usual exposure to an educational experience did not bring about the results desired by society, it was hypothesized that an additional exposure would compensate or put the learner back on grade level. Thus compensatory education was a suggested technique at both the national and State levels.

In its broadest outlook compensatory education was designed to supplement, not supplant, facets of ongoing education. In New York State, for example, learners who are not up to grade level continue to be taught the regular reading material that is paid for by general aid. Moreover, such learners are taught an additional set of lessons paid for by compensatory aid. Thus, the exposure to reading for some pupils is specialized, more concentrated, and provided for a longer period of time. Devising a means for offering such supplemental work has centered in the "project" concept. The word project as used here means a planned action involving the learning of a phase of school work resulting in the general achievement by the student of some specifically stated objectives.

In the 1972-73 academic year, there were 218 funded projects in the Urban Education Program in New York State. While a detailed breakdown of the total expenditures will be presented later in this section, it is important to know at this point that there were forty-seven million dollars[1] available for projects in the Urban Education Program. The remainder of the report will concern itself with the specifics of such projects.

c. Procedures Used in the Urban Education Program - How Urban Education Projects are Initiated, Approved, Financed and Amended

Urban Education funds are allocated by the Legislature on an annual basis. They are then divided among urban school districts in New York State on the basis of a predetermined formula. This formula establishes the maximum grant available for an individual district.

---

[1] Less $180,000 ($90,000 each) in the form of a direct grant to Roosevelt and Wyandanch, rather than in the form of reimbursement for approved expenditures.

After a district is notified of its Urban Education maximum grant allocation and the current guidelines and priorities of the Urban Education program, it then submits a project proposal or several project proposals for reimbursement of approved local expenditures. The district has the option of submitting a proposal for a completely new project or for the continuation of the previous year's project. In theory, projects which meet desirable objectives are resubmitted in subsequent years. Additionally, projects which don't quite meet their objectives are improved through changes in methodology or objectives and are also resubmitted.

Project proposals submitted by local school districts to the Urban Education Office fall into two categories, new and resubmitted.

New Project Proposals are reviewed by the professional staff member responsible for the district submitting the proposal. The Urban Education staff then determine which other units within the Department should review the proposal. The proposal is then routed to the designated subject area specialists for review of the educational approach and to the evaluation unit for review of the proposed evaluation design.

Each unit involved recommends approval or disapproval of the proposed expenditures to the professional staff member in the Urban Education office. The Urban Education staff member then coordinates the various reviews. The final decision for approval or disapproval is made by the responsible staff member in the Urban Education office. After the project'sproposed expenditures have been determined to be reimbursable by the Urban Education office it is sent to the state-aided programs unit in the Department for a review of the project budget in terms of appropriateness and accuracy. The revised/approved budget is sent to the local district directly by the state-aided program unit.

Resubmitted project proposals are for projects that have been approved and in operation in previous years. They are reviewed for approval only by the responsible professional staff member in the Urban Education office and the evaluation unit.

Recycled projects need not be reviewed by subject area specialists. They are sent a copy of the proposal. They may review the project proposal but it is not required.

The evaluation unit makes its recommendation for approval or disapproval of the recycled project's evaluation design to the Urban Education staff member. The staff member then makes the final decision for approval or disapproval.

As with new projects, after the recycled project has received final approval it is then sent to the state-aided programs unit of educational finance for review and approval of the budget.

Urban Education funds allocated by the Legislature on an annual basis are suballocated by a predetermined formula which takes into consideration (1) number of children from families receiving AFDC, (2) number of sixth grade pupils scoring below level on the PEP test in 1967 and (3) weighted average daily attendance of the district. It should be noted that, since 1967 data continues to be used rather than more recent results, each district's proportionate share of available funds has remained constant.

Each district eligible for Urban Aid gets a proportion of the total allocation based on the formula. The formula allocation for each school district receiving Urban Education funds in 1972-73 is presented in Table 18. For purposes of comparison of the magnitude and distribution of these funds, the 1972-73 allocations for various other compensatory aid programs are also listed in this table.

Amendments to project proposals fall into three categories: program change or modification, budget change, or both program and budget change.

The amendments are sent to the state-aided programs unit. If the modification includes both program and budget changes the program office reviews the modification and then returns the approved program modification to the state-aided program unit. Amendments containing budget modifications are accepted for processing by the state-aided programs unit in September, December and March. Only

Table 18. Approved Funding By Source in Urban Education Districts 1972-73

District	Urban Education 1/	School Community Intersection Umbrella	E S E A				Pre-K	Total
			Title I	Title II	Title III	Title VIB		
New York City	$39,098,556	$1,998,153	$137,818,673	$2,536,289	$1,765,051	$914,629	$2,821,524	$186,952,875
Buffalo	2,385,823		7,100,030	226,002		116,843		9,828,698
Rochester	1,183,819		4,451,740	135,439	50,761	181,475	220,255	6,223,489
Syracuse	630,298		2,871,317	74,141		201,751	402,924	4,180,431
Yonkers	612,429		1,710,353	45,856	275,000	57,387	195,443	2,896,468
Niagara Falls	400,335		1,076,505	29,665		25,136		1,531,641
Schenectady	292,968		622,395	20,397		51,872	223,560	1,211,192
Albany	245,179		1,483,794	24,395	116,332	47,776		1,917,476
Mt. Vernon	230,590		1,099,765	20,916		14,186	94,015	1,459,472
Utica	169,868		981,760	53,150		37,779		1,242,557
Elmira	145,737		457,510	53,338				656,585
New Rochelle	120,795		361,394	12,752			313,333	808,274
Rome	113,849		492,753	13,965		12,530		633,097
Hempstead	93,380		729,953	8,934			242,806	1,075,073
Troy	89,954		548,507	50,199			57,221	745,881
Freeport	78,271		616,576	8,656		20,651	194,407	918,561
Poughkeepsie	53,322		376,728	8,998			293,894	732,942
White Plains	43,125		234,418	6,442			113,047	397,032
Long Beach	42,166		512,398	8,075			103,841	666,480
Amsterdam	38,476		202,881	6,374			145,484	393,215
Port Chester	36,221		352,041	7,274				395,736
Middletown	33,305		195,760	9,039			116,571	354,675
Jamestown	31,030		289,668	7,800				328,498
Lackawanna	30,687		290,081	5,108				325,876
Fulton	28,977		182,857	6,612				218,446
Glen Cove	23,789		189,621	7,100			106,392	326,902
Ossining	19,765		187,524	3,030			96,459	306,778
Binghamton	16,251		567,527	31,475		26,384	88,114	729,751
Watertown	12,737		2/ 230,921	8,455	96,628			117,820
TOTAL	$46,301,702	$1,998,153	$166,004,529	$3,429,876	$2,303,772	$1,708,399	$5,829,290	$227,575,721

1/ Wyandanch and Roosevelt public schools receive a flat grant of $90,000 each for a total of $180,000, under Section 3602, Subdivision 11, paragraph g of Chapter 553 of the Laws of 1973.

2/ The exact value was not available at the time that the totals were computed.

one modification per project is accepted during each of of these periods.

Comparative evaluation is a continuing effort to obtain information for making decisions. Additionally, the best decisions are based on cumulative results. Thus, evaluation findings can and should be used to bring about changes in a project's objectives or activities during the actual life of the project.

Minor modifications may be made without prior consent, but major changes in program, objectives, evaluation design, or budget require such approval.

Some types of project modifications are shown in Table 19.

From the table it may be seen that a relatively small proportion of projects had modifications in objectives or activities prior to or after project implementation. Additionally, the majority of such modifications took place after consultation with the State Education Department (SED).

Table 19

Urban Education Project Modifications

Modification	No. of Projects Indicating		Percent of Total Indicating	
	Yes	No	Yes	No
Objectives prior to implementation	12	206	5.5	94.5
Approved by SED	6	6	50.0	50.0
Objectives after implementation	5	213	2.3	97.7
Approved by SED	5	0	100.0	0.0
Activities prior to implementation	11	207	5.0	95.0
Activities After implementation	22	196	10.1	89.9

d. A Profile of Urban Education Projects

Data in this section were collected from all Urban Education school districts on mailed report forms. Satisfactory information was obtained from all districts.

Community Involvement in Planning

School districts are urged to elicit the help of district residents in establishing Urban Education programs. Such help may be in the form of planning, needs assessment or the establishment of priorities. Since sheer numbers in a given category do not indicate community participation, several sets of data were generated to obtain a broader view of this aspect of Urban Education programs.

From Table 20 it may be seen that the District Advisory Committee was the group with the largest number of participants and also was the type of group most frequently used by school districts for project planning. Approximately 80 percent of the projects utilized 3 or more of these groups in their planning efforts.

Table 20

Number of Individuals Who Participated in Planning at the District Level

	Number	
	Participants	Projects
District School Board	1,316	170
District Advisory Committee	6,142	209
Local PTA or Home/School Association	5,844	132
Representatives from Local Community Organizations	1,337	151
Nonpublic School Officials	499	107
Other Groups	1,352	72
TOTAL	16,490	

Project Time Span

Projects could fall into one of three broad time categories as shown in Table 21.

Table 21

Project Time Span

Time Span	Number of Projects	Percent of Total
Regular School Year	178	81.7 %
Summer Only	9	4.1
Calendar Year (12 months)	31	14.2
Total	218	100

From the table it may be seen that the larger number of projects were those which were funded for the regular school year. Similarly, 4.1 percent were funded for the summer only and 14.2 percent were funded for a full calendar year or for a 12-month period.

Table 22 presents a more detailed analysis of project time span for the most common components found in Urban Education projects. This table presents information on the average number of weeks and hours per week that critical activities in these project components take place.

Table 22

Average Duration of Urban Education Offerings

Component	Activity	Number of Weeks		Hours Per Week	
		Activity Operates	Individual Participants	Activity Operates	Individual Participates
Reading Grades 4-6	Diagnostic & Remedial	35	31	23	8
Reading Grades 1-3	Diagnostic & Remedial	35	34	23	10
Reading Grades 7-9	Diagnostic & Remedial	36	31	21	6
English as a Second Lang. Grades 10-12	Develop-mental	38	38	38	4
Reading Grades 7-9	Develop-mental	38	34	22	12

The table is read as follows: If a component is operated 40 weeks for three hrs. a day, and 6 groups of pupils are each scheduled for 1 hour's instruction daily for 20 weeks, columns 2-5 in the table (reading from left to right) would be 40; 20; 15; 5.

Another aspect of project time span is the number of years a project has been in existence. From Table 23 it may be seen that of the total projects in operation, the majority (59.2 percent) were operational for three or more years.

Table 23

Years of Project Operation

Years	Number of Projects	Percent of Total	Cumulative Percent
First Year	36	16.5 %	16.5%
Second Year	53	24.3	40.8
Third Year	68	31.2	72.0
Fourth Year	41	18.8	90.8
Five or more	20	9.2	100.0
Total	218	100.0	

Of the 218 projects, 63.8 percent were expected to be resubmitted in the 1973-74 academic year, 2.8 percent were not expected to be resubmitted, and 33.4 percent had an uncertain future according to appropriate district personnel. The Division of Urban Education had analyzed 1972-73 projects in relation to the reading, mathematics, and bilingual education priorities to be mandated in 1973-74. Districts were then advised about the potential for recycling 1972-73 projects in 1973-74 in relation to the priorities.

Enrollment by Ethnic Group and Project Type

District coordinators of Urban Education projects were requested to use a standard procedure for assessing student ethnic backgrounds and to provide the number in each ethnic group, who were expected to participate as indicated in the proposal as well as the number of actual participants.

From Table 24 it may be seen that the largest number of participants were Black. Information from 23 projects was classified as "failed to designate".

Table 24

Ethnic Group

Ethnic Group	Participants				Number of Projects	Failed to Designate		Total Actual Partic.
	Expected	Percent of Total	Actual	Percent Of Exp.		Actual No. of Partic.	Expected Number of Projects	
American Indian	303	0.1%	177	58.4%		4		181
Black	190,979	41.9	166,527	87.2		7,592		174,119
Oriental	4,098	0.9	3,640	88.8		65		3,705
Spanish Surnamed American	201,897	44.3	165,177	81.8		4,280		169,457
Other Including White	58,167	12.8	52,511	90.3		5,332		57,843
TOTAL	455,444	100.0	388,032	85.2	195	17,273	23	405,305

From Table 25 it may be seen that the anticipated participants in both the public and nonpublic schools exceeded the actual participants by 9.7 percent in the most common components of Urban Education projects. Additionally, nonpublic school participants constituted less than one percent of the total actual participants for these components. Participation of nonpublic school pupils in regular, in-school programs is illegal. However, as district residents nonpublic school pupils are eligible for off-hours, week-end, and summer Urban Education activities. However, throughout the State, the total number of nonpublic school participants in Urban Education projects was reported as 62.543, the latter represents 15.4 percent of the total participants in the 218 Urban Education projects.

Table 25

Participants in Five Major Components
by Type of School

Component	Activity	Objective	Type of Participants					
			Public School Participants		Nonpublic School Participants		Total	
			Expected	Participating	Expected	Participating	Expected	Participating
Reading Gr. 4-6	Diagn. & Reme.	Achieve.	14,724	13,367	10	8	14,734	13,375
Reading Gr. 1-3	Diagn. & Reme.	Achieve.	5,948	5,716	20	20	5,968	5,736
Reading Gr. 7-9	Diagn. & Reme.	Achieve.	10,086	9,738	5	5	10,091	9,743
ESL	Develop.	Achieve.	6,659	5,053			6,654	5,053
Reading Gr. 7-9	Develop.	Achieve.	7,763	7,298			7,763	7,298

Table 26

A Comparison of Actual Participation by Project Component between the Urban Education Program and ESEA, Title I Program

Project Component	Urban Educ. Program Actual	ESEA, Title I Program Actual
Supportive Services	120,446	41,362
Reading	110,038	272,076
Cultural Enrichment	73,592	30,013
Mathematics	52,053	118,657
Other Basic Skills (Science, Social Studies, English, Foreign Languages)	50,530	30,137
Pupil Personnel Services	42,164	49,436
Inservice Education	26,235	11,107
Curriculum Development	16,724	2,579
English as a Second Language	13,077	15,491
Health Education	7,539	7,710
Bilingual Education	6,554	16,335
Vocational Occupational Education	5,652	1,696
Adult Career Education	3,884	490
Handicapped	1,365	8,649
TOTAL:	529,853	605,851

A statewide profile of the priority emphasis given to different aspects of education in Urban Education projects can be inferred from Table 26. These priorities are also compared with a profile of E. S. E. A. - Title I compensatory education projects in Table 26. The first survey of all components of Urban Education projects revealed that Supportive Services touched the largest number of pupils (120,446) in the Urban Education Program and that reading treatments involved the largest number (272,076) of E. S. E. A. - Title I pupils (Table 26). These data reflect the differences in program purposes. The Urban Education program in 1972-73 was continuing in supporting the concept of community education activities including programs for out-of-school youth and adults. The second largest number of component-participants in the Urban Education Program were enrolled in reading activities, while the second largest number of component-participants in E. S. E. A. - Title I were in mathematics. Mathematics ranked fourth for Urban Education component-participants.

A comparison of absolute numbers of component-participants is not especially meaningful since (1) the E. S. E. A. - Title I allocation is over $3\frac{1}{2}$ times that of Urban Education, and (2) the different program offices use different recommended per pupil expenditures (not in a ratio of 3.5:1), because of the difference in funding levels and program emphasis.

Staffing Patterns in Urban Education Projects

The number of elementary and secondary level staff who were directly engaged in project activities are shown in Table 27. The secondary level staff are divided into two categories. "Basic Skills" involve fundamental subjects such as reading, writing, mathematics, natural and physical science and social science. "Vocational Skills and Attitudes" relate to areas such as business education, home economics, industrial arts and other occupationally oriented instruction. For each type of staff, both the full-time equivalence[1] and the unduplicated number of persons are shown. All direct project staff are reported in this item. From the table it may be seen that the largest single number of staff employed were paraprofessionals at the elementary school level.

---

1 For the completion of the item, the total "full-time equivalent" of staff members was calculated. Each staff member was counted as a whole of a fraction of the equivalent of an individual who works full time.

Table 27

Number of Direct Staff Employed by Level and Type
in Urban Education Project Activities

Type of Staff	Elementary Level Staff		Secondary Level Staff			
			Total			
	Total Full-time Equivalent	Unduplicated Number	Total Full-time Equivalent	Unduplicated Number	Total Full-time Equivalent	Unduplicated Number
Teachers	838.7	1194	760.0	1378	1598.7	2572
Other Professionals	123.6	153	86.1	137	209.7	290
Paraprofessionals	1628.8	3321	764.2	1615	2393.0	4936
Nonprofessionals	78.5	123	201.8	164	280.3	287
Total	2669.6	4791	1812.1	3294	4481.7	8085

The number of staff providing supportive services for each project was not included in Table 27. "Supportive Services" staff are those providing auxiliary services to the project either to participating pupils or in some other way, but who are not directly engaged in project activities. There were about 700 individuals who provided these services. Adding these to the staff listed in Table 27 we see that in order to implement the 218 Urban Education projects, approximately 8,800 individuals were employed. It is also worth noting that in 187, or 85% of these projects, over 90% of the teachers employed were certified or licensed.

Every project activity is expected to have one licensed or certified staff member. In addition, some teachers may have an aide or an assistant who may provide direct, non-professional staff support services. Such additional staff are referred to as "paraprofessionals".

The types of services provided by paraprofessionals may be gleaned from Table 28 while in Table 29 are shown the number of projects using paraprofessional services.

Table 28

Number of Urban Education Projects
Employing Paraprofessionals

Mode of Employment	Number of Projects
Not employed	43
Noninstructional classroom duties	32
Tutoring	87
Small group instruction	133
Other	71

Responses were received from 218 of the projects funded and of these the greatest number of projects employed paraprofessionals for small group instruction while 37.6 percent of the projects used paraprofessionals for noninstructional classroom duties.

Additionally, 43 projects (19.7 percent) did not employ paraprofessionals. It may be noted that 61 projects (28 percent) employed paraprofessionals for three or more different services.

Table 29

Number of Types of Paraprofessional Services Used in Urban Education

Number of Services	Number of Projects	Percent of Total
None	43	19.7 %
One	49	22.5
Two	65	29.8
Three	50	22.9
Four	11	5.1
TOTAL	218	100.0

Component Activity Costs

The components of all projects were listed. Additionally, the appropriate major activities were related to the respective components. Finally, a dollar amount associated with each component and activity was summed for the entire state. From Table 30 it may be seen that the single largest component was reading at the elementary school level which used diagnostic and remedial activities. Table 30 also includes estimates of the average cost per participant hour of these activities. Data of these types are critical for good decision making in the area of internal resource allocation among programs. For example, the question of at what level of education diagnostic

and remedial reading should be offered obviously depends on at what level such services are most productive in terms of student benefits. But in cases where productivity is relatively constant at different grade levels, this cost information provides some indication of where it can be offered most cost-effectively.

Table 30

Component Activity Costs for the Component Activity Combination with the Five Highest Expenditures

Component	Activity	Number of Projects	Total Cost	Average Cost per participant hour
Reading Grades 4-6	Diagnostic & Remedial	32	$3,195,069	$0.84
Reading Grades 1-3	Diagnostic & Remedial	23	2,300,646	0.83
Reading Grades 7-9	Diagnostic & Remedial	25	1,380,116	0.64
English as a Second Lang. Grades 10-12	Developmental	2	1,368,103	1.78
Reading Grades 7-9	Developmental	9	1,057,292	0.31

e. Resource Distribution by Program Types

In the school year 1972-73 $227,575,721 of categorical aid was budgeted for use in local school districts. This figure includes $46,301,702 for the Urban Education Program. The remainder of the funds was allocated under the following other categorical aid programs: E.S.E.A. Titles I, II, III, VI-B, the school community Interaction Umbrella, and the Experimental Prekindergarten programs.

Table 31 indicates how local school districts utilized Urban Education as compared to the "other" programs identified above in terms of project components. The data indicate that both the Urban Education and the "other" programs emphasized the three priority areas of the Department - Reading, Mathematics and Bilingual Education. The Department expects that the percentages in these categories will be even higher in 1973-74 because local school districts will have had more time to readjust their programs to the priorities.

Table 31

Comparison of Percentages of State Approved Budgets by Project Component for Categorically Aided Programs in Urban Education Districts 1972-73

Project Component	Urban Education	Other Sources	Combined Sources
Reading	29.0 %	37.7%	36.0%
Administration	10.6	4.3	5.2
Mathematics	10.1	15.2	14.2
Supportive Services and Others	10.0	4.0	2.8
Pupil Personnel Services	8.2	7.8	3.4
English as a Second Language	7.2	3.5	.5
Other Basic Skills (Science, Social Studies, English, and Foreign Languages)	4.8	9.6	8.7
Bilingual Education	3.7	6.8	6.1
Inservice Education	3.3	1.6	7.9
Cultural Enrichment	2.8	1.5	.5
Handicapped	2.6	3.6	1.8
Evaluation	2.5	.0	5.5
Adult Career Education	1.7	.2	.3
Curriculum Development	1.7	.0	.5
Early Childhood	.9	3.2	1.9
Vocational Occupational Education	.8	.4	4.3
Health Education	.1	.6	.4
TOTAL	100.	100.	100.

A large percent budgeted for Urban Education was Administration which was nearly $2\frac{1}{2}$ times higher than it was (on a percent basis) for all other sources. Recalling, however, that the Urban Education budgets represent only 20 percent of all categorical aid budgets discussed here and that the problem of purchasing administrative services has a minimum cost for a project which tends to decrease as a percent, relative to a rise in participant enrollment (with a corresponding rise in direct instructional services), the high percent budgeted to administration may be a function of the total budget rather than a trend toward swollen administrative staffs. The fourth and fifth largest budgeted components are "Supportive Services and Others" (e.g., food, transportation of library aides),

and "Pupil Personnel Services" (e.g. guidance, psychological), which combine to 18 percent in Urban Education as compared to 12 percent on the average for the other categorical aid sources. The fifth largest budgeted component for Urban Education was "English as a Second Language" (serving largely Spanish language dominant youngsters). Urban Education budgeted over twice the percentage amount that the other categorical programs budgeted for helping disadvantaged non-English speaking learners adjust and function quickly in their new English language dominant society.

Urban Education, then, was similar to the other categorical programs in that reading and mathematics (combined) received the largest portions of the budgeted resources. Urban Education budgets did reveal a difference in program emphasis from the other programs in that a greater percentage was budgeted for pupil personnel services, other support services, and English as a Second Language, in line with its programmatic emphasis.

In terms of the combined totals for the Urban Education and "other" programs the data indicate that reading received 36 percent of all budgeted monies, mathematics 14 percent, other basic skills 9 percent, and Bilingual Education 6 percent, for a combined total of 65 percent.

Resource Distribution Within Programs

Table 32 presents a breakdown of the major categories of budgeted expense within Urban Education programs. Similar information is also provided for other categorical aid efforts (enumerated above for purposes of comparison).

The information displayed in Table 32 reveals that the category "Personal Service-Instructional" across all categorical programs discussed there involved 35 percent ($79,624,052) of the total. Fifty-five percent of the Urban Education budgeted funds were allocated to the categories "Personal Services-Instruction", whereas 34 percent of budgeted funds from Other Sources were

allocated to the same categories. The percents of budgeted funds allocated to "Personal Services-Noninstructional: and "All Others" was 15 from Urban Education and 44 from Other Sources. With respect to staff costs, these comparisons indicate that Urban Education approved budgets were oriented more toward supporting instructional costs, while noninstructional and other costs were financed to a higher extent in the approved budgets of projects funded from Other Sources.

Table 32

Percent of State Approved Budgets,[a/] by Budget Category, For Categorically Aided Programs in Urban Education Districts, 1972-73

Budget Categories	Urban Education	Other Sources	Combined Sources
Personal Services-Instruction	39.2	33.9	35.0
Other Instruction	16.0	.1	3.3
Employee Benefits	10.0	12.7	12.1
Administration	8.7	9.1	9.0
Personal Services-Noninstruction	8.5	23.7	20.7
All Others	6.6	10.4	9.6
Equipment, Supply	6.0	4.7	4.9
Evaluation	2.4	1.6	1.8
Cont. Agree., Cons. Services	1.4	.9	1.0
Trans. of Pupils	.7	.9	.9
Conf. and Inservice	.4	2.0	1.7
Parent Involvement	.1	.0	.0 *

[a/] Includes approved budgets available for evaluation and analysis as of October 1, 1973

* Rounding concealed the $25,920 (less than .1 percent) budgeted for parent involvement found in Urban Education budgets.

f. Results - The Achievement of Students in 1972-73 Urban Education Programs

This section of the report contains a statistical analysis of data reported to the Department by school personnel for each Urban Education project.

Departmental staff reviewed the data supplied by local districts. Of the 218 projects funded during the 1972-73 program year, standardized test data from 114 projects were usable in the statistical analysis. Data for more than one group of participants were reported in many projects; in all, there were 469 groups. Groups were distinguished for such reasons as multiple components within a project, different schools participating in a project, or participants from the four grade intervals[1] under investigation. The number of sampled participants in the groups ranged from 3 to over 2,000. To assure a greater reliability in the pretest and posttest means, groups containing fewer than ten participants were eliminated from the statistical analysis.

How Progress Was Determined, and by Whom

According to approved evaluation plans, project personnel administered a standardized test prior to the beginning of the project treatment (e.g. a reading program) and the same test (usually an alternate form) at some specified point during or at the end of the project treatment. The first test administration is the pretest, and the second is the posttest. For some groups the pretest was administered as much as five months prior to the beginning of the project treatment. For example, a project component conducted in a community district of the New York City Public Schools may have specified the measurement of pupil achievement by using reading scores obtained from the Metropolitan Achievement Test administered during the citywide testing program

---

[1] The four intervals studied were early primary (grades 1-3); later primary (grades 4-6); early secondary (grades 7-9); and later secondary (grades 10-12). Analysis by intervals permits the inclusion of nongraded treatments without extensive data manipulation by pupil chronological age/grade placement.

which occurs in April of each year. If the project component did not commence until September, then there was a five-month lapse between the pretest administration and the initiation of the treatment. Additionally, if the project lasted for the entire ten months of the school year, then only seven months of the treatment would be reflected in the posttest score. Thus, any difference score, i.e. posttest score minus pretest score, is confounded by two limitations: (1) the uncertainty of the effect of multiple experiences between the pretest and the initiation of the program, and (2) a diminution of the net gain reported due to lack of congruence between the testing dates and the dates between which instruction took place. Large-scale field evaluation commonly suffers from limited exercise of the controls specified in most experimental designs. Therefore, caution must be exercised when deriving conclusions from the data and the results of the data analysis.

Departmental specifications required districts to transform participant raw scores to grade-equivalent scores and compute the mean grade-equivalent for each group on both the pretest and the posttest. District staff were required to determine whether the difference between the predicted and actual posttest means was statistically significant for each group. During the review of the data, Departmental professionals concluded that statistical analyses of the data should be limited to tests conducted upon group pretest, predicted posttest, and actual posttest means as the most reliable sources of data reported by district personnel.

The Department established a mathematical procedure for district staff to employ in predicting each participant's pretest grade equivalent score based on the number of months the participant had been in school prior to the pretest, and the number of months which elapsed between the pretest and posttest. District staff calculated and reported the pretest mean for each group. Using these means, Departmental staff derived a predicted mean gain and an actual mean gain for each group. The predicted mean gain is defined as the difference between the predicted posttest mean and the actual pretest mean. The actual mean gain is defined as the difference between the actual posttest mean and the pretest mean. These mean gains were

divided by the number of months between the administration of the pretest and the posttest to obtain a monthly mean gain and a monthly actual mean gain for each group. These two statistics form the input data for the analyses reported in this section and show if the project actually could have speeded up students' rate of learning.

During 1972-73, local education agencies spent up to 5 percent of their allocations to evaluate the effectiveness of the funded projects. The Department devised a plan for assessing the effectiveness of the project but some districts, especially those in New York City, expressed a desire to obtain more information. Some districts employed a contractor to undertake the necessary evaluative work. The types of evaluators selected by districts are exhibited in Table 33.

Districts outside of New York City (NYC) made very little use of evaluation contractors (that is, commercially-oriented agencies). Rather, school district personnel usually completed the required evaluation work with the aid of a consultant.

Table 33

Types of Evaluators

Type of Evaluator	Urban Education	
	Number	Percent
District Personnel (Outside NYC)	48	22.0 %
NYC Bureau of Educational Research (BER)	6	2.8
NYC School Staff (Other than BER)	7	3.2
University, firm, etc.	157	72.0
Total	218	100.0

Achievement Progress by Students in the Urban Education Program

In this section of the report, the record of achievement of the Urban Education Program is studied by viewing 459 subgroups of pupils for which data were available. These subgroups were obtained from the total project groups by subdivision in accordance with subject studied and grade level. Use of the subgroups, or project components, is particularly appropriate because the districts reported their data as arithmetic means of the subgroups.

Pupils were evaluated in two ways: (1) actual gain, meaning the gain achieved during the period of participation in an Urban Education project; and (2) in terms of predicted gain, meaning the gain estimated for the same period of time under regular instruction. This estimate is based on rate of achievement prior to entering the Urban Education Program. The means of these two measures were then reported to the Education Department by the subgroups.

On the basis of the two measures, it may be said that the operation of urban aid has resulted in improvement for a subgroup if the actual gain mean exceeds the predicted gain mean and as lacking in improvement if the converse is true. In addition, the difference--mean actual gain minus mean predicted gain--provides a measure of improvement. The attached table and charts display the data obtained from this point of view.

In Table 34 are shown total numbers of subgroups and associated numbers of these subgroups for which improvement has occurred. The succeeding column provides corresponding percents. It can be seen that there has been improvement for 78 percent of all subgroups, 89 percent of subgroups in grades 4-6 reading, and so on.

Table 34

Number and Percent of Project Component Subgroups Showing Improvement by Subject Matter of Instruction

Subject Taught	Number of Subgroups		Percent Improved
	Total	Improved	
Basic Skills	40	27	68 %
Reading:			
Grades 2 - 3	53	44	83
Grades 4 - 6	112	100	89
Grades 7 - 9	78	58	74
Grades 10-12	9	9	100
Other	12	12	100
Mathematics	94	65	69
Other Subjects	61	41	67
All Subjects	459	356	78

NOTE: Improvement - A project is considered as improved if actual gain exceeds predicted gain

In order to obtain a comprehensive measure of participant performance regardless of component treatment, a weighted mear was calculated by multiplying each group's mean by its number of participants, summing all such results, and dividing by the number of participants. Weighted means were calculated for each component and for all components, the results of which appear in Table 35. The overall results showed that the predicted monthly gain from pretest to posttest was 0.61 of a month, i.e., approximately 13 days during a 21-day school month. The corresponding actual monthly gain from pretest to posttest was 0.94, i.e., approximately 18 days during a 21-day school month. The actual gain surpassed the predicted gain by approximately 50 percent. Additionally, the actual gain achieved by the

Table 35

Number of Participants and Weighted Mean
by Project Component Category

Project Component	Total Number of Participants	Number of Sampled Participants	Weighted Monthly Mean Gain	
			Predicted	Actual
Mathematics	8,744	3,952	0.61	0.90
Reading	61,927	25,035	0.56	0.93
Other Basic Academic Skills	10,343	3,224	0.80	1.02
Other Than Basic Academic Skills	27,375	3,958	0.68	0.94
All Components	108,389	36,169	0.61	0.94

sampled participants is approximately that which is expected of all school pupils, 21 days in a 21-day school month. Because of the nature of the evaluation design, which lacked comparable control groups, the nearly-normal gain cannot be attributed explicitly and solely to treatments funded under the Urban Education Program. Many factors contributed to the gain attained by the sampled participants, but this group had at least one common factor which was a program treatment funded under the provisions of the New York State Urban Education Program.

5. Testing to Measure the Effectiveness of Bilingual Education

The Department is faced with a different set of problems than those discussed above when the area of bilingual education testing is examined.

The Position Paper on Bilingual Education of the Board of Regents of the University of the State of New York highlights the education problems of an increasing number of public school students whose dominant languages are not English. While priorities and strategies for bilingual program develop-

ment and classroom instruction have been disseminated, and school districts and teacher education institutions have a variety of bilingual education activities underway, evaluation of the effectiveness of bilingual education programs is in a rudimentary stage.

Recognizing this problem, the Department formed a Bilingual Education Evaluation Committee which functions in concert with a Statewide Advisory Committee on Bilingual Education. Participating in the evaluation committee are representatives of the Puerto Rican Educators Association, the Office of Bilingual Education - Board of Education, New York City, several higher education institutions involved in bilingual teacher education and research, and local district program directors, teachers and psychologists participating in operating bilingual education programs.

The Bilingual Evaluation Committee fulfilled its initial mission by identifying the following measurement needs in order of priority:

a. Development of instrumentation which could serve to measure the effect of specified bilingual programs on student achievement. The recommendation pointed out that the greatest need is to measure the degree to which linguistic skills taught to students, whose major or only language is Spanish, are being successful in equipping the student to read, write, understand and speak in English.

b. Development of valid tests which could measure student achievement in content areas such as mathematics, social studies, science, etc., when instruction is given in Spanish.

In addition to these priorities a third suggested priority of the State Education Department was to evaluate a sample of bilingual and English as a Second Language programs which had a promise of program success in advance of

the instrumentation project (a above), which itself would not have any products for field testing until school year 1974-75.

The following is a report on the progress made in three projects designed to meet the previously stated three priorities. These projects are functioning as components of the state evaluation system operations in 1973-74 and the following descriptions account for their first four months' operations from September through December 31, 1973.

a. Spanish-English Linguistic Skills Test Development

The first activity of this project involved researching what were the greatest needs in the area of Spanish-English linguistics measurement. A determination was made that instrumentation was most needed to measure the degree to which students 1) could comprehend written and spoken language in both Spanish and English; 2) could write in both languages; and, 3) could speak English and Spanish. The second problem identified for study was whether or not existing instrumentation developed primarily for Mexican-American bilingual students was linguistically transferable or adaptable for use by students whose Spanish background reflected the Caribbean area.

After these initial specific recommendations were formulated, review and development work was designed. Up to the time of this report, a collection of teacher-made and standardized and non-standardized tests developed for linguistic measurement was made and reviewed, and a series of conferences held with teachers and project directors of bilingual educational projects to determine the degree of instructional emphasis on each of the four linguistic areas pointed out above. Based upon a summary of these preliminary investigations, it will be determined whether new instrumentation developed or reviewed is needed by operating projects for measurement in all the linguistic areas mentioned above.

The most widely used tests in bilingual programs operating in the State are the Inter-American series in both Spanish and English, the Metropolitan Achievement Test and the Wide Range Achievement Test. Other tests reviewed include

those developed by Burt, Spolsky-Murphy, Cervenka, and Pimsleur, all of which are in the experimental stage. It is planned that certain tests used in the schools of Puerto Rico as well as additional teacher-made tests will also be examined. This aspect of the bilingual testing operation is proceeding as planned and on schedule.

A continuation of this and the following test administration project in 1974-75 calls for State administration and coordination of testing in categorically aided Bilingual and English as a Second Language projects whose objectives will be measurable by the instrumentation developed. Activities for 1975-76 will include the development of alternate forms of the final tests selected, the refinement of test items, and the development of student achievement proficiency criteria.

b. State Bilingual Testing

As mentioned previously, a second interim effort is being undertaken to evaluate bilingual education projects with existing evaluation procedures in advance of the development of new measurement devices.

As an initial step, all Urban Education funded programs designated as either bilingual programs with instruction both in Spanish and English, and "English as a Second Language" programs with instruction only in English skills for foreign language dominant students were reviewed in detail by the Department. Eight projects were identified as operating in eight local districts whose proposals had evidence of sound instructional practices and whose timetable of implementation appeared realistic. An additional criterion of selection was that the best possible evaluation design should be included as part of the project. The districts were New York City Community Districts 23, 24, 20, 10, 13, 11; the City District of Long Beach on Long Island, and the Westchester County City District of Yonkers.

In these eight projects field visits were made to insure that the evaluation design and instructional program would be as strong and sophisticated as possible. Arrangements have been made to acquire all pre and post test data adjudged to be the best available to measure the types of objectives and skills taught in each of the projects for State evaluation purposes. In many cases, recommendations have been made

for the substitution of new tests and instruments for those originally planned, in order to strengthen evaluation components.

A report on this activity will be reviewed early this spring.

c. Testing of Science Skills when Instruction has Been in Spanish

Although it was determined that the greatest priority need in terms of programs was in the development of an instrument in the Spanish language which would test those social studies and citizenship education skills which had been taught in Spanish, the problems of measuring attainment of those skills and attitudes as specified objectives in the New York State courses of study for social studies indicated that initial attempts at measurement development should be in another subject matter field. Science education was selected because it appeared that reasonable progress could be made in developing a new test. Grades 1 through 3 have been identified as those which operated more bilingual education project components with science instruction in a foreign language. The area of Spanish was selected, since it was the language of bilingual instruction for the vast majority of students in such programs.

After deciding on the area (i.e. science) an analysis of the State syllabus in science for grades 1 through 3 was commenced. In addition, a classification system by the content of each specified topic of the required study units was begun. A further classification of the learning outcomes expected of the students, as stated in the syllabus for each unit, was followed by a plan for test item writing. Recommendations from these activities were then reviewed and plans finalized for developing suitable tests.

In the coming months teachers in bilingual programs who are teaching science in the Spanish language in grades 1 through 3 in New York State will be contacted and asked to submit single copies of teacher-made tests and any other tests available which would measure the various content topics and skills as determined in the syllabus unit classification system. The work plan calls for the validation and refinement of teacher-made test items and the supplementing of any content topics in the syllabus in which good

teacher test items cannot be located, by production of new test items by project consultants. The product is to be a series of question test items which correspond to all of the topics in the State science syllabus, grades 1 through 3, and a report of the results of a tryout of these science test items by a panel of teachers in Spanish science instruction scheduled for 1974-75.

As of the date of this report 118 teachers in New York State have been identified as being involved in such programs and as having a high degree of participation and interest. It is now anticipated that science tests can be assembled by classroom teachers, administered, and results reported back during the program year 1974-75. Tentative Department plans in 1974-75 call for the initiating of a project to develop similar test items in social studies for the same grades and according to the same methodology described above in the 1974-75 year, with field testing slated for 1975-76. The products produced will be adapted to any changes made in the elementary science course of study.

B. QUESTION #2: WHAT TYPES OF INSTRUCTIONAL PROGRAMS (OR PROJECTS IN THE CASE OF CATEGORICAL AID PROGRAMS) ARE MOST EFFECTIVE FOR STUDENTS OF DIFFERENT CHARACTERISTICS?

In Chapter III - the projected program - the point is made that the absence of adequate computer software capability has made it virtually impossible for most educators to trace what really happens in instruction. In that Chapter, the description of the Instructional Evaluation System (IES) indicated how this barrier can be eliminated in the future. In the interim program described in this Chapter the Department indicates how it is collecting less aggregated information than has been possible in the past but it must be noted that analyses described here are not as sound as those that will be possible when the Instructional Evaluation System is implemented in all districts. Two additional

points should be made: first, whenever aggregate information is used, for example information describing an instructional project funded under the State's Urban Education Program, one is apt to find as much variation in what actually is implemented between classrooms and buildings as is found between projects; and, second, there are some professionals who question some of the statistical approaches used. On this latter point, the Department will undertake a complete review of these approaches this spring.

The description of the following five work efforts deals with question #2 above: 1) New York City Prototype Study of Urban Education and E.S.E.A. - Title I Programs - School Year 1973-74; 2) Evaluation of Urban Education Projects - School Year 1972-73; 3) Urban Education Performance Indicators Study - School Year 1972-73; 4) Characteristics of High Performing Upstate Schools; and 5) Study of the Effect of Pupil Mobility Upon Pupil Evaluation Program (PEP) Results.

1. New York City Prototype Study of Urban Education and Elementary and Secondary Education (E.S.E.A.) - Title I Programs - School Year 1973-74

   a. Purposes of the Study

      The purpose of this prototype study is to test the feasibility of developing and implementing a Department (as opposed to district) managed evaluation program which would also be designed to further the development of the projected program described in Chapter III. New York City community districts were chosen for the study because of the large percentage of funds which they receive from compensatory aid programs (i.e. Urban Education and E.S.E.A. - Title I).

      It was determined that present local evaluation efforts although useful for local purposes prohibit meaningful comparison across districts in that each district is free to use whatever test(s) it desires, and administer tests at different times of the year. In addition, little is done at the local

level in examining the relationship between students receiving services under categorical aid programs and those not receiving those services. This study deals with these two barriers and definitely is yielding more comparable data available for decision making. The evidence presented on pages 48-60 shows, based on the community district sample of all grade five students in reading that questions regarding unmet need (i.e. the number of students needing services as evidenced by pretest scores list not receiving services from categorical aid funds) is now defined. Several questions regarding whether funds are allocated to students with the greatest need can now be addressed with data.

b. Procedures Used in the Study

A standardized test of reading comprehension (the California Test Bureau - McGraw-Hill Comprehensive Test of Basic Skills Form Q - Level I) was administered as a pre-test to all grade five students in 10 New York City community districts in early November of 1973. In May of this year, a post-test will be administered. In addition, an experimental version of the "reading effectiveness measure" will also be administered in the spring of this year. The administration of this test will assist the Department in refining its testing procedures for the "projected" measure. It will determine whether by holding the difficulty of test items consistent for a passage with a certain "readability", student responses follow a set pattern so it will be possible to more precisely indicate his or her ability to comprehend different passages with varying "readability" scores. The scaling of the passages for "readability" is not as precise as that desired in the ultimate measure but is an improvement over what has been feasible in the past.

In addition to these measures of output, an improved set of information regarding each project has been collected. Specifically, a program questionnaire was completed by each Urban Education and E.S.E.A. - Title I coordinator in each of the ten community districts for grade five reading projects. Twenty-eight questions were used to examine the following variables which are thought to have an impact on outcomes of projects: 1) how well the project objectives and instructional procedures are articulated; 2) how much of a change in instructional and organizational procedure does the project introduce; 3) how individualized is the project; 4) how time intensive is the project (i.e. how many

hours are the children exposed to instruction); 5) how labor intensive is the project; 6) how well managed is the project; and 7) how well is the project documented. The classroom teachers involved with the project were asked for additional information. Specifically they were asked to provide information on the reading textbooks by reading group, the social study textbooks used in grade five, an indication of the number of students in each ethnic group, and some general information regarding their experience. Additional information on the teachers is available from the BEDS file. It should be noted that the purpose for collecting all of this information is for analysis of projects, not individual students or teachers. The State is only interested in determining to the best of its ability what are exemplary projects and in attempting to identify the characteristics of projects which are worth replicating.

The information collected in this study is being stored on computerized student and program files. This is being done to make it practical to undertake analyses after the post-testing is completed. A description of the types of analyses to be performed which address question #2 follows a brief discussion of what the Department is providing districts in this study.

Feedback of Information to Participating Districts

In undertaking this prototype study, the Department determined that it should feed back as much information to the participating community districts as possible. Generally evaluative and descriptive information is provided to State and federal agencies but little information is provided for the school districts. As an example of how the Department is attempting to reverse this practice, after the pretest scores were computed by the Department and entered on the computerized student record by teacher, printouts such as Tables 36 and 37 were distributed to each classroom for the teacher's use. In addition, similar tables were provided to principals and the superintendent. After the post-testing similar tables will also be forwarded to the same individuals.

The participating community districts are also very interested in receiving back information from the "experimental version of the reading effectiveness measure". As indicated previously, each classroom teacher indicated what reading and social studies textbooks were being used. The

Table 36

Individual Skills in Reading Comprehension

Percent of Right Answers for Comprehensive Test of Basic Skills, Reading Comprehension (Level 1)

District _____ School _____ Teacher _____ Page _____

Superintendent _____ Principal _____ Study # _____ Date __/__/__

Student (1)	Sound Recognition(2)	Literal Meaning(2)	Main Idea(2)	Relation-ships(2)	Conclu-sions(2)	Total (3)
Name #						
Name #						
Name #						
Name #						
Total						

(1) Test not valid because teacher reported: Student Ill (I), English too new a language (L), Other (O).

(2) The questions on which this analysis is based were answered correctly at least 50% of the time by a large national sample of students in the sixth month of grade four.

(3) Based on the 35 test items referenced in footnote 2.

Table 37

Grade Equivalent, National Percentile, and District Percentile Based Upon Comprehensive Test of Basic Skills, Reading Comprehension (Grade 5.2)[1]

District _____ School _____ Teacher _____ Page _____

Superintendent _____ Principal _____ Study # _____ Date __/__/__

Student[2]	National Grade Equivalent	National Percentile[3]	District Percentile[3]	Student[2]	National Grade Equivalent	National Percentile[3]	District Percentile[3]
Name #				Name #			
Name #				Name #			
Name #				Name #			
Name #				Name #			

[1] All results are based on all 45 test items.

[2] Test not valid because teacher reported: Student Ill (I), English too new a language (L), Other (O).

[3] Derived by RRI. Because the Reading Comprehension Test has only 45 items, some percentile values do not occur.

"readability" of these textbooks are now being calculated. Their "readability" level will be compared with the standardized pretest scores of students to determine whether or not the grade-level difficulties of the textbooks correspond to the grade level ability of the students. In addition, after the "experimental version of the reading effectiveness" measure is administered, the Department believes that it will be possible to determine whether or not students have met the reading standards set by their reading and social studies texts. Information from this work will be very helpful to districts because generally they individually have not previously undertaken this type of effort.

c. Planned Types of Analyses - Spring and Summer 1974

As mentioned previously, in the spring of 1974 (May and June), two posttests are planned--1) a standardized posttest in reading comprehension will be given; and, 2) an experimental version of an effectiveness measure for reading will be administered to determine if students are able to comprehend English text at a level of difficulty corresponding to the expectations implied by the assignment of particular reading materials. Growth in reading comprehension from November, 1973 to June, 1974 will be evaluated relative to the actuarial growth in the comprehension skills of students in various percentile bands measured from data obtained from a national norming sample.

Many other analyses of the data can be executed. These potential analyses are of three general types. Each type is discussed in subsequent paragraphs.

Since students in compensatory programs are not ordinarily distinguished from students who are in regular programs when standardized tests are normed, statistics such as grade equivalents, percentiles, etc. already incorporate the gains (if any) produced by categorically-funded compensatory programs. Thus, for example, the differences among students revealed by standardized test statistics may be smaller than they would be if disadvantaged students were not participants in categorically-funded programs at the time a test was normed. This confounding effect would serve to reduce the likelihood of detecting any program effects with the tests. In addition, for numerous other reasons, there are major methodological problems associated with using a national sample (or an urban sample) as a comparative basis for evaluating the benefits of Urban Aid programs.

However, the problems associated with evaluating program effects in relationship to the distribution of test scores obtained from a national sample can be avoided to an appreciable extent if a program versus no-program (experimental versus control) design could be employed. In this type of analysis, those students at each level (percentile band) of pretest achievement who are taking the program could be compared with matched students who are not in the program.

The problem is to determine whether designs of this type are feasible in the 1973-74 evaluation. That is, it will be necessary to review the pre-test data and the program documentation that have been provided by the pilot districts in November and December to determine whether there are enough comparable students taking and not taking compensatory programs. It is the Department's minimum expectation that it may be possible to draw subsamples of students matched on pretest scores and other variables (for example, use of volunteers, all of comparable age, sex, and equal proportions of comparable ethnic backgrounds) from the available student populations (i.e. those in and out of programs). Because of the desirability of this type of analysis, the Department is now reviewing the pretest results, information concerning who is in which program, how students were selected, and the program documentation itself in order that a decision to use program-by-no-program (i.e., experimental versus control) designs can be made rationally.

Another type of analysis that is possible involves the comparison of schools within districts or the comparison of programs within districts (or both). Such analyses would be valuable to determine whether one program was more effective than another. In addition, on the hypothesis that one program proved to be more effective than another, one might wish to determine, for example, whether the differential application of resources accounted for the result. However, such analyses depend upon determining whether the programs are run under comparable conditions. Therefore, before a decision is made to undertake analyses of this type, their feasibility must be determined by examining the ethnic distributions of classes among schools, the program documentation, pretest scores, and so forth.

A third type of analysis is also possible. This type involves comparisons between districts based on such variables as staff-student turnover, the quality of programs as determined by staff analysis, etc.. Again, a large amount of information is being reviewed in order to determine whether such analyses are feasible before a decision can be made concerning their execution.

In summary, the pilot approach to Department-managed evaluation is as follows. Give a standardized pre- and posttest in reading comprehension and a post-test measure of effectiveness in reading. Obtain information concerning pupil-program documentation. Execute the following analyses: 1) a calculation of the actual growth of students in each program and an evaluation of the results relative to the growth expected on the basis of actuarial data obtained from a national norming sample; and 2) a calculation of the extent to which students are able to comprehend text equivalent in difficulty to that which they are using in instruction. Finally, display all data (e.g., frequencies, means, variances, etc.) for reporting purposes, and review the results in order to determine the feasibility of: program versus no program designs; within-district program comparisons; and the feasibility of using both BEDS-derived indicators and the results of a program standards evaluation for making between-district comparisons. (It is apparent that analyses of pre- and post-test data can only be undertaken for students who took the pre- and posttests. Thus, all evaluation studies will be based only upon analyses of test scores of students who were in the program long enough to profit from it.)

Obviously, there is a wealth of information available for analyses. The decisions regarding which specific analyses are possible and are of a higher priority will be made shortly.

2. Evaluation of Urban Education Projects - School Year 1972-73

On pages 60-86 , the rationale for the Urban Education Program, descriptive information and results associated with the expenditures were reported. In the following pages, the specific "exemplary projects" are summarized

since they are the projects which have been proven to be successful. The compendium contains data on project goals, activities and accomplishments and is intended to serve as a guide to local personnel and to answer question #2 - What types of projects are most effective for students of different characteristics?

Selected abstracts of projects are presented which meet the following criteria: To qualify, an exemplary project is one which equalled or exceeded one month gain for each month of instruction, contained valid supporting data, had reasonable per pupil expenditures and were funded in some of the priority areas.

a. Reading Laboratory Program

The project was operated in New York City Community District 2 with a budget of $85,910 and 307 participants in grades 4 through 7 (cost/participant $280).

Major Goal: To increase reading skills of elementary and junior high school students who utilize a reading laboratory as measured by the Metropolitan Achievement Reading Test.

Activities: Reading laboratories were established in one elementary and two junior high schools to provide an individualized self-directing, and self-correcting approach to the elimination of reading difficulties. Students worked in the laboratory in groups of fifteen with each group scheduled for two periods of work each week. Each student was diagnosed through testing, and recommendations were formulated based on the student's needs as evidenced in the diagnosis. A schedule of work was prescribed for each student. The program made extensive use of innovative and programmed materials such as workbooks, language arts games and reading skill tapes and employed filmstrips, teaching supplies, spoken arts cassette library, controlled readers, tape recorders, listening stations and earphones.

Findings: Student reading gains were predicted to be at six months but gains were actually achieved at a rate of one year and one month. At the fifth grade level the rate of growth reached one year and nine months while at the sixth grade the rate of growth reached one year and six months. These gains were all significant at the .01 level of confidence. For the program as a whole, 65 percent of the pupils improved their rate of growth by 50 percent or more.

b. Operation Reading Success

The project was operated in New York City Community District 19 with a budget of $748,021 and 1,502 participants in grades 4 through 6 (cost/participant $499).

Major Goal: To diagnose and correct reading deficiencies to achieve a growth of one year in one year.

Activities: Each school participating in Operation Reading Success was equipped with a reading laboratory staffed by a teacher and two educational assistants. Each laboratory was designed to diagnose and correct specific reading deficiencies in a small group setting. Participants attended the laboratory for one hour a day. A class of 16 pupils, grouped by ability, was in attendance at any one time. There were four such classes in each of 19 schools. The laboratory included three stations composed of a listening station, an Aud-x station, and a combination of a controlled reader and TACH-X stations. Each station was designed to correct specific reading deficiencies and improve, stimulate and motivate the participating child.

Findings: The mean pretest score for the group was 3.09 and the posttest score was 4.01. Sixth grade students achieved at a rate of one year and one month. The average rate of growth historically was four months while the actual growth was significantly higher at the .001 level.

c. Junior High Reading Laboratories

The project was operated in New York City Community District 22 with a budget of $161,056 and 473 participants in grades 6 through 9 (cost/participant $340).

Major Goal: To improve reading ability of pupils who are one or more years behind so that they are able to return to regular classes.

Activities: Each laboratory was supervised by a reading teacher who was aided by two educational assistants. The students were programmed for a half year with five double periods or five single periods a week. Those students who did not achieve a singificant gain in a half year remained for a full year. In September each pupil was given a pretest with the Metropolitan Achievement Test in Reading. A diagnosis was made of each student's needs using the Gilmore Oral Reading Test and the Roswell Chall Phonic Test. Every student's eyes were tested with the Titmus Vision Tester. A program was planned individually for each student based on his deficiencies. Students worked individually or in small groups. The teacher and assistants taught and gave help as necessary. After testing, students returned to their English classes if recommended by the teacher.

Findings: Ninth grade students attained one year and one month's growth, eighth grade students attained one year and four months' growth while seventh grade students attained nine months' growth. Predicted growth was four months and all groups exceeded this rate at the .001 level of confidence.

d. Diagnostic Reading Program

The program was operated in New York City Community District 27 with a budget of $559,972 and 1,589 participants in grades 2 through 9 (cost/participant $352).

Major Goal: To improve the reading level of students in grades 2 through 9 after diagnosis and specific skill instruction.

Activities: Second grade children who scored below grade 1.5 on the Metropolitan Achievement Test were grouped according to similar needs for small group instruction. They received instruction in listening skills, auditory and visual discrimination and practice in decoding and comprehension skills utilizing the Flashcard Reader, cassette player and listening centers. In the Reading Helper aspect of the program upper level children reading at least six months below grade level as measured by the Stanford

Diagnostic Reading Test were given supplementary instruction in reading skills by paraprofessionals. The Reading Helpers worked with a maximum of three children at a time who were given specific help. Each child was screened for visual deficiencies using the Keystone Telebinocular.

Findings: Achievement gains for the eight group averaged one year as measured by the Metropolitan Achievement Test. Achievement increase at the fifth grade was one year and four months and at the sixth grade it was one year and seven months, both significant at the .01 level of confidence.

e. Cluster Team Program

The project was operated in New York City Community District 31 with a budget of $157,008 and 500 participants in grades 6 through 8 (cost/participant $314).

Activities: The program was designed to provide a learning environment and instructional methods for students with serious reading handicaps. The children in the program were economically as well as educationally deprived. Pupils in grades 6 through 8 met at least twice a week with the reading teacher. A teacher assistant provided small group instruction. The instructional method employed multi-media and multi-sensory equipment and techniques. The student first received perceptual accuracy and visual efficiency training followed by activities which were intended to enrich his experiential background and prepare him for subsequent instruction. The student then developed vocabulary, word recognition, word attack skills, and comprehension skills. Finally, all of the words, skills, and concepts taught during the skill-building sequence were applied by the student during independent reading. Achievement was measured by use of the Metropolitan Achievement Test.

Findings: The average achievement increase for grades 6-8 was one year and three months.

f. Home Study Center for Disadvantaged Children

The project was operated in Binghamton City Schools with a budget of $20,190 and 393 participants in grades 1 through 12 (cost/participant $51).

Major Goals: To provide tutorial assistance for needy students, to diagnose special problems, to provide individual help and with parents to encourage students to do better in school.

Activities: A minimum of one hour of individual tutoring per week was provided in the home where possible and when not possible at a Homework Center. Tutors were either local high school students, college students or volunteer adults from the community. The homework centers were equipped with appropriate materials for study and an atmosphere conducive to study. Liaison was established with school officials and faculty for purpose of mutual cooperation and resolution of academic problems of tutees. Children in grades 1 through 12 were served. The staff-participant ratio in project activities was 1 to 1.

Findings: The improvement in reading and mathematics was one year and one month which was significantly greater than the historical rate of growth for the students.

g. Prescribed Reading Essentials Program

The project was operated in Ossining Public Schools with a budget of $19,727 and 150 students in grades 7 through 9 (cost/participant $131).

Major Goals: To improve reading vocabulary and comprehension as measured by the California Reading Test after a program stressing individualized instruction.

Activities: Teachers conducted reading classes ranging in size from 1 to 10 students. Classes were occasionally conducted for the groups but usually students worked individually after having been appropriately programmed according to their special needs and levels. The teachers conferred with parents of participants and a building-parent group was involved in the process. The teacher aide supervised students working on individual tasks and occasionally helped the teacher organize classroom materials. The target group was ninth graders reading below level after which other students were invited to participate. Students were scheduled for reading two or three times a week.

Findings: The average achievement gain was one year and two months.

h. Uplifting of Basic Skills

The project was operated in the City School District of Troy with a budget of $90,000 and 500 participants in grades 1 through 6 (cost/participant $180).

Major Goal: To improve skills in reading, mathematics, science and social studies through a program which stresses reading skills, increased interest in reading, resource materials which children can use for independent study and teachers can use in their classrooms and by providing a resource teacher to aid and motivate children in independent study.

Activities: Various instructional methodologies and program materials were applied across the group. In one case, forty-five children in grades four through six were helped by a remedial reading teacher. Grouping in this instance was based on individual weaknesses rather than by grade with emphasis placed on comprehension, phonics, word building and vocabulary. In another part of the program, thirty children who had not been achieving up to their capacity worked in a large room which could be divided for small group instruction. Team teaching and individualized instruction were used for teaching reading and mathematics with emphasis on work skills, vocabulary and computational skills. The overall district program also included a library-instructional media resource center which provided materials for remedial reading components.

Findings: A sample of 141 students was tested using the Metropolitan Achievement Tests for grades 4-6. Achievement gains were found to be one year and two months and results were statistically significant at the .02 level of confidence.

i. Secondary Education-Reading Programs

The project was operated in the Utica Public Schools with a budget of $165,568 and 965 participants in grades 7 through 12 (cost/participant $176)

Major Goal: To develop reading skills and promote positive attitudes toward reading.

Activities: At the high school level a reading center was operated in laboratory format with a reading teacher and assistant. The reading teacher was responsible for diagnosing reading skill needs and for planning, developing and implementing instructional activities to improve those skills. The assistant helped students in the use of equipment and materials, provided tutorial help and maintained records and materials. The effectiveness of the center in the improvement of reading was measured with the Stanford Diagnostic Reading Test and the Stanford Achievement Test.

In the junior high school two types of reading centers were used. One type consisted of an after-school program utilizing student tutors. Students were scheduled for three one-hour sessions per week for 15 weeks. The aim of each session was to improve reading skills through positive tutor student interaction and varied learner activities utilizing multi-media and multi-level approaches. The other type center was a part of the regular day program which emphasized group instruction.

Findings: Average achievement gain at the ninth grade level was one year and five months. This is statistically significant at the .01 level of confidence.

j. Community School

The project was operated in White Plains Public Schools with a budget of $43,241 and 112 participants in grades 8 through 12 (cost/participant $386)

Major Goal: To carry out an alternative educational program at the secondary level for students not reached by existing secondary school programs.

Activities: Students participated in academic work three days each week and small group career orientation for the other two days. The staff consisted of a director, a community liaison coordinator, four full-time and five part-time academic teachers, a guidance counselor, a part-time social worker, and fifty-two volunteer teachers who provided community courses. The school offered a full range of academic offerings as well as field experience courses directed toward career and self-exploration. Close personal contact was provided between teacher and students in math and reading instruction. Students were given

prompt assistance with their work, as well as frequent consultations with teachers in the diagnostic and remedial phases of the program. Readily available manipulative aids were provided in mathematics.

Findings: Students achieved at a rate of 1.4 months for each month of participation in the program.

In addition to Urban Education funds, all Urban Education Districts receive ESEA Title I funds to provide assistance for educationally disadvantaged pupils. Some ESEA Title I projects in these Urban Education districts were selected as exemplary. The criteria for selection of these projects was the same as those previously described for Urban Education.

k. Bilingual Program

The project was operated in New York City Community District 3 with a budget of $145,857 and 425 participants in grades K through 3 (cost/participant $343).

Major Goal: To demonstrate a significantly different mean on the posttest measure of Spanish reading than on the pretest.

Activities: Eighteen bilingual classes were organized in a language arts curriculum. Reading content included a portion of cultural heritage material, an action that had strong endorsement by parents. The project operated for 40 weeks at 25 hours per week. Each participant received 520 hours of supplementary instruction. Seventeen teachers and 22 paraprofessionals were involved.

Findings: The Cutler Diagnostic Spanish Reading Test consists of 5 subtests, all of which use raw scores as measures of accuracy. For each subtest, the posttest mean was found to be significantly different from the pretest. Thus, the data led the evaluator to conclude that target population had achieved considerable growth in ability to read in Spanish between the time of the pretest and posttest.

l. Mathematics Laboratories

The project was operated in New York City Community School District 31 with a budget of $223,846 and 357 participants in grades three, four, five, and seven (cost/participant $627).

Major Goal: To achieve beyond expectation in mathematics concepts and computation as measured by the Stanford Diagnostic Arithmetic Test.

Activities: A diagnostic analysis of pupil deficiencies was followed by prescriptive activities planned by the laboratory staff and classroom teachers. The physical manipulation of materials was emphasized for concrete experience prior to abstractions. The Mathematics Laboratories were conducted for 36 weeks for 35 hours per week. Individuals participated 5 hours per week for 36 weeks which provided 180 manhours of math instruction supplement per pupil. The equivalent of 9 full-time teachers and 8 nonprofessionals were employed in the treatment.

Findings: The target population averaged 18 months growth in achievement in a 7-month period between the October pretest and May posttest administration of the Stanford Diagnostic Test. Growth expected without the ESEA, Title I treatment ranged from $3\frac{1}{2}$ to 6 months.

m. Services to Early Elementary Children

The project was operated in New York City Community District 15 with a budget of $1,403,789 and 2,599 participants in grades Kindergarten through 2 (cost/participant $540).

Major Goal: To achieve a significatn difference in reading between pretest and posttest means as measured by the Metropolitan Reading Readiness Test.

Activities: A programmed instruction approach using the DISTAR method was used. Three hundred forty reading lessons were given. The project operated 36 weeks for 30 hours per week. Each individual received 180 hours supplementary treatment. The project employed 46 teachers and 157 aides.

Findings: Random samples (50%) from each grade achieved significant differences between pretest and posttest means. In second grade, California Achievement Test scores indicated a mean of 9 months growth in the 7 months between pre and posttest administrations.

n. Bilingual Program

The project was operated in New York City Community District 19 (Brooklyn) with a budget of $1,010,212 and 2,600 participants in grades 1 through 8 (cost/ participant $389).

Major Goal: To achieve a significant improvement in Spanish reading as measured by the Cutler Diagnostic Test in Spanish.

Activities: Participants received English as a Second Language instruction alternately with reading instruction in Spanish. The project operated 40 weeks for 32 hours a week, employing 59 teachers and 12 aides. Participants received 1,040 hours supplementary treatment.

Findings: A sample of 303 students pretested in January, 1973 yielded an average raw score of 73.9 from a possible 112 points. The post test mean of 97.1 obtained in June, 1973 was found to be significantly different at the .01 level.

o. Reading Diagnostic Center

The project was operated in New York City Community District 29 (Queens) with a budget of $312,605 and 625 participants in grades 1 through 4 (cost/participant $500).

Major Goal: To achieve in reading at a rate of 1 month growth for each month of participation as measured by the Spache Diagnostic Reading Test.

Activities: A clinical approach with psychological and reading diagnosis was followed by prescription of skills related to code breaking. The project operated 40 weeks for 30 hours a week. The participants received 200 hours of supplementary instruction. The staff included 9 teachers, 9 educational assistants, 1 psychologist, 2 social workers, and 1 family assistant.

Findings: Among the participants for which both pretests and posttests were available, 40 participants in grade 1 achieved 10 months in the 7 months between tests, 195 grade 2 participants achieved 14 months; 274 third graders achieved 14 months, and 35 grade 4 participants achieved 13 months.

p. English as a Second Language in Nonpublic Schools

The project was operated by the New York City Central Board with a budget of $634,392 and 3,976 participants in grades 1 through 12 (cost/participant $160).

Major Goal: To significantly improve English language oral fluency by pupils with limited or no English speaking competency as measured by the New York City Board of Education Pupil Ability Scale.

Activities: Direct instruction was provided in listening comprehension and oral facility. Materials included tapes, filmstrips, phono viewers, as appropriate to age-interest levels. The project operated 40 weeks for 5 hours per week. Each pupil received 200 hours of supplementary instruction. The staff was composed of 42 teachers.

Findings: The mean rating of a sample (372 participants) stratified by grade rose from 5.836 to 8.043 on the scale. The scale consists of a rating schedule of 1 to 6 points.

q. Remedial Reading Clinic

The project was operated in Buffalo with a budget of $147,338 and 423 participants in grades 3 through 6 (cost/participant $348).

Major Goal: By the conclusion of the project year, all students participating were expected to demonstrate a significantly greater rate of growth in reading during the 10-month period of the program than they had demonstrated in the past on the Reading subtest of the Metropolitan Achievement Test.

Activities: A clinical team observed and tested pupils. This provided diagnosis and prescription for a treatment carried out through a small group or tutorial setting in twelve target area schools. The treatment ran 40 weeks for 15 hours per week. Individual participants received 120 hours of supplementary instruction. The staff included one reading specialist, 16 reading teachers, 16 school aides, 4 reading diagnosticians, and 4 psychologists.

Findings: A sample of 423 participants who were expected to achieve $4\frac{1}{2}$ months' growth in ten months actually achieved 14 months' growth in this period as measured by the Metropolitan Achievement Test (reading).

r. Corrective Mathematics

The project was operated in Buffalo with a budget of $541,821 and 2,058 participants in grades 2 through 8 (cost/participant $263).

Major Goal: To demonstrate a significantly greater rate of growth during the 10-month period of the program than had been demonstrated by students previously measured by the Metropolitan Achievement Test.

Activities: A tutorial approach and a laboratory approach were employed with different participants. The most severely disadvantaged pupils participated in the lab sessions, while participants in the 2nd and 3rd stanines received tutoring. The treatment lasted 37 weeks for $27\frac{1}{2}$ hours per week with the learner receiving 111 hours supplementary math instruction. Thirty-seven math teachers, 27 teacher aides, and 27 support pupil personnel service staff were employed in the project.

Findings: Two hundred forty-two pupils, who were predicted to achieve 5 months' growth, participated in the math labs. These pupils actually attained 14 months' growth as measured by the combined Metropolitan Achievement Test. The 1,082 disadvantaged learners in the tutorial sessions, who were predicted to attain 5 months without tutoring, achieved 15 learning months as a result of the program.

3. Urban Education Performance Indicators Study - School Year 1972-73

a. Purpose of the Study

For several years, the Department has been developing the Performance Indicators in Education (P.I.E.) project as a statistical method of determining how various characteristics of the district relate to achievement. Generally, the data used in the statistical model has been BEDS (Basic Education Data System) information and PEP (Pupil Evaluation Program) scores. As part of the evaluation for the present year, this work is being extended to examine specific Urban Education projects. The purpose is to develop Urban Education performance indicators to determine how various characteristics of projects relate to achievement on a project rather than on a district basis.

b. Procedures to be Used in the Study

Three plans of analyses have been developed to investigate the following: 1) the feasibility of studying reading and arithmetic achievement by simultaneously developing equations for both criteria; and, 2) the effects of program components, activities, and objectives; of program duration and of pretest scores on achievement.

The analyses are based on data on compensatory aid projects as submitted to the Department in the Mailed Information Reports (a data collection instrument used by the Department to collect descriptive and evaluative information from local school districts) for the 1972-73 school year. Selected data for 80 projects have been extracted from MIR's and are being analyzed. Analysis of additional data for these 80 projects and of data for other compensatory projects is dependent upon the availability of data being assembled by the Bureau of Urban and Community Programs Evaluation.

c. Anticipated Results of the Study

1) Information concerning the effects of program components, activities, and objectives and the interaction of these factors. For example, the results might show that the objective of improving reading achievement has been achieved better through activities A, B, and C than through activities X, Y, and Z in a population with given characteristics.

2) Information concerning the effect of program duration on student achievement.

3) Information concerning the relation of pretest scores to achievement.

In the coming years, this type of study will be expanded to cover each school year with the explicit purpose of determining those project characteristics which have different effects on students under various conditions in different projects. As a result of these analyses, unmeaningful items can be deleted from the MIR forms and dissemination of information regarding successful projects can be improved.

4. Study of the Characteristics of High Performing Upstate Schools

a. Purpose of the Study

The purpose of this component is to identify characteristics of upstate schools which are highly successful in promoting student achievement given the conditions under which they operate.

b. Procedures to be Used in the Study

A project is presently being finalized with the Bureau of Institutional Development of SUNY at Buffalo. That organization already has received the following data through a separate arrangement with the Information Center on Education: 1) Basic Educational Data Systems file for 1971 for all schools in the State; 2) Pupil Evaluation Program files for 1969-1972; 3) Financial File for all districts; and 4) Personnel File for all districts. The Bureau of Institutional Development is presently merging these files.

This work complements on the school building level activities of the Bureau of School Programs Evaluation to develop performance indicators on the school district level.

Specific objectives of the project are:

1) To identify schools which are achieving above expected levels in elementary school reading and arithmetic.

2) To identify those school and community characteristics which are related to varying school performances in reading and arithmetic.

The Department is using funds to refine an observational instrument which was used in a study of urban schools, and to test it in a number of upstate schools. Proposals have been solicited from potential contractors and consultants.

c. Anticipated Results of the Study

The major anticipated outcomes are:

1) Information on which BEDS data items which can account for a degree of school performance.

2) A list of school management and instructional practices to be examined further for their relation to student achievement.

3) An observational instrument designed to examine management and instructional activities in nonurban schools.

4) A plan for a study in depth of student, community, and program factors related to school performance.

This study will be expanded in fiscal year 1974-75 to identify school management and instructional practices related to school performance in upstate schools. This rationale for this work effort is to verify judgments on activities and conditions which promote student achievement since they are presently confounded by the many interrelationships between school factors, nonschool factors, and student characteristics. The Performance Indicators project is an effort to separate these factors so that those which can be affected by school personnel, primarily the school program factors, can be examined to determine how achievement can best be enhanced. The study underway and proposed for 1974-75 carries this process further by examining a number of factors at the school, rather than on the district level.

This stream of inquiry is designed to make statewide Performance Indicators more useful as a school management tool. In 1975-76, this project would be phased out as a separate endeavor and its findings integrated into the Urban Education Performance Indicators stream, as described above.

5. Study of the Effect of Pupil Mobility Upon Pupil Evaluation Program (PEP) Results

a. Purposes of the Study

Each year, after the PEP test results have been reported to the schools, it is not uncommon to have principals complain that because they have a large pupil turnover, they are being unfairly "blamed" for the educational deficiencies of pupils who have been in their schools for only relatively short periods of time.

Of course, what these principals fail to realize is that it is not at all a matter of "blame". The primary purpose of the PEP program is to identify and locate a target group of pupils for special educational programs. If pupils are properly classified in the target group, on the basis that their scores fall below the Statewide Reference Point established for that purpose, how long the pupils have been enrolled in the school is not a particularly important consideration for purposes of educational planning. The point remains that students in the target group have certain educational needs, and the number of target group pupils in a school is an index of unmet needs. The validity of this index is in no way diminished by the fact that more or less of the pupils are transferees.

Nevertheless, it is true that there has been a tendency in some sectors of the press and the public to equate the size of the target group in a particular school with the effectiveness of the educational program. Certainly that error would be compounded if pupil mobility were ignored. It was determined, then, that it would be useful to explore the dimensions of the mobility problem, and the relationship between pupil mobility and PEP scores.

b. Procedures Used in the Study

To provide a picture of pupil mobility, a sample of public schools participating in the regular October 1972 PEP testing program were requested to provide two separate summaries

of reading test scores for Grade 3 pupils, as follows: 1) all pupils enrolled, including transfer students, and 2) only those students who had been continuously enrolled in that school since beginning Grade 1.

The sample was drawn on a stratified basis. In New York City, the sample consisted of all schools in Local Districts 6, 17, and 24, selected because their ethnic composition and October 1971 test scores were generally typical of New York City as a whole. Of the 7 large city districts in the State, 6 were included in the sample; and of the 8 medium-size city districts, 7 were included. In the small-city, village-suburban, and rural community types, roughly 10% of the school districts in each type were included, again selected on the basis of the criterion of representativeness in the October 1971 testing. Table 38 indicates the number of districts, school buildings, and pupils in each community type in the sample.

Table 38: Districts, School Buildings, and Pupils in the Sample

Community Type	School Districts	School Buildings	Pupils
New York City	3	43	7,004
Large City	6	204	13,963
Other Cities	13	127	7,916
Village-Suburban	25	80	8,031
Rural	38	55	3,609
Total	85	509	40,523

c. Results of the Study

Table 39 indicates that transfer students are a common phenomenon in the schools of the State, even as early as Grade 3. The highest rate of transfer students was found in the sample of New York City schools, where a total of 37% of the Grade 3 pupils had enrolled initially in Grade 1 in some other school. The lowest rate was found in the rural schools, but even here the rate of transfer

among Grade 3 pupils was 19%. When the various community types are weighted for relative enrollments in the State, the transfer rate for the State as a whole in Grade 3 may be estimated at 29%.

Table 39: Relative Frequency of Third Grade Transfer Pupils

% of Transfer Pupils	NYC Schools	NYC %	Large Cities Schools	Large Cities %	Other Cities Schools	Other Cities %	Village Suburban Schools	Village Suburban %	Rural Schools	Rural %
Over 50%	2	5%	21	10%	15	12%	2	2%	3	5%
49-40	13	30	27	13	10	8	4	5	1	2
39-30	17	40	49	24	21	16	12	[illegible]	1	2
29-20	9	21	59	29	38	30	28	35	16	[illegible]
19-10	1	2	37	18	38	30	27	34	27	[illegible]
Below 10	1	2	11	6	5	4	7	9	7	13
Total	43	100%	204	100%	127	100%	80	100%	55	100%
% of Transfer Students in Combined Group	37%		34%		26%		26%		19%	

However, it must be kept in mind that the transfer rate varies rather widely among individual schools. Some schools in each community type have third grades made up of 50% or more of transfer students, while other schools have third grades made up of fewer than 10% of transfer students.

That pupil mobility is indeed related to pupil achievement is suggested by a comparison of the reading test results of transfer and non-transfer pupils (Table 40). In New York City, for example, whereas 57% of the transfer students in Grade 3 had scores below the Statewide Reference Point for identifying target group pupils, the corresponding target group rate for non-transfer students was only 36%. Outside of New York City, however, the relationship was less marked. It would, therefore, appear that the problem of transfer

students varies with the type of school. Outside of New York City, for example, transfer students may generally be coming from schools with educational programs comparable to that of the new school, whereas in New York City the transfer students may be more likely to be educationally disadvantaged in comparison with non-transfer students in the same school.

Table 40

Comparative Test Results for Transfer and Non-Transfer Students

	Percent Target Group Pupils Below Statewide Reference Point			
Community Type	Total Enrollment	Non-Transfer Pupils	Transfer Pupils	Difference: Transfer Minus Non-Transfer
New York City	44%	36%	57%	21%
Large Cities	36	33	43	10
Other Cities	21	19	26	7
Village-Suburban	14	13	17	4
Rural	18	17	22	5
Statewide Estimate	26%	21%	36%	15%

Granted that the reading test scores of transfer students in a school may tend to be lower than the scores of non-transfer students, the question still remains whether this factor has a significant effect on the comparison of any particular school with other schools in the same community type. That is, assuming that all transfer students were systematically excluded from PEP test reports, it might be expected that the target group rates in all schools would tend to be lowered. But if the degree of change in target group rate is fairly constant, then the inter-school comparisons within a particular type would remain relatively high when transfer students are eliminated, and schools with unusually low target group rates would remain relatively low.

To determine what the practical effect would be if transfer students were eliminated from PEP test reports, the correlation was found for each community type between the target group rate in the total enrollment and the target group rate among non-transfer students only. Table 41 indicates that among schools in large cities, for example, the correlation was .96. That means that, with reference to the percentage of target group pupils, the relative standing of any particular school among all other schools in large cities was virtually the same whether transfer students were included or excluded. Even within the same large city, as well, the relative standing remained unchanged. Further, very much the same picture was found for New York City schools and for schools in all other community types.

Table 41

Correlation Between Percent of Target Group Pupils in Total Enrollment and Percent of Target Group Among Non-Transfer Pupils

Community Type	School Buildings	Correlation (Pearson r)
New York City	43	.93
N. Y. C. District 6	(11)	(.93)
N. Y. C. District 17	(13)	(.92)
N. Y. C. District 24	(19)	(.91)
Large Cities	204	.96
Albany	(17)	(.95)
Buffalo	(68)	(.97)
Rochester	(40)	(.91)
Syracuse	(31)	(.95)
Utica	(19)	(.98)
Yonkers	(29)	(.98)
Other Cities	127	.94
Village-Suburban	80	.95
Rural	55	.94

d. Summary of the Findings

The frequency of transfer students, defined as students who began enrollment in Grade 1 in another school, is fairly high even by the time that the students reach Grade 3. In the State as a whole, it is estimated that about 29% of the Grade 3 students tested in the October 1972 PEP program were transfer students. As a group, the transfer students did in fact tend to have lower PEP reading test scores than the non-transfer students. However, this student mobility factor did not seem to have a sufficiently differential effect on the PEP results of schools in the same community type to affect the relative standing of individual schools. That is, if transfer students had been eliminated from the PEP reports, the relative rank order of individual schools in Grade 3 reading scores, as compared with other schools of the same community type, would have remained virtually unchanged.

III. DESCRIPTION OF PROGRESS IN DEVELOPING THE PROJECTED PROGRAM: A COMPREHENSIVE STUDENT EVALUATION PROGRAM

The Department's "projected" program described in this Chapter has been developed over the past three years. Many factors were considered. First, questions such as those outlined in the introduction were formulated. Second, needs of decision-makers at various levels and interested parties such as the Legislature, parents and the general public were examined in the context of what information was needed to improve decision-making, and to improve reporting to parents and the public. Third, and most important, the answers to all questions were pursued with a view to improving the educational system so that students would receive a better education. Fourth, reviews were made of existing systems used in this State and elsewhere, specific analytic research and development activities undertaken, and pilot tests were implemented to test out new approaches thought to be suitable. In addition, two "state of the art" reviews were seen as being crucial to developing this "projected" program. The reviews of (1) existing testing methodologies, and (2) availability of disaggregated information are presented under "A" below.

All of the work was conducted within what is technically labeled as a "systems approach". The Department quickly realized that evaluation of instructional programs could not be isolated from other aspects of the instructional process such as instructional program planning, instructional management, or the development of an information system able to provide "feedback" data on the results of implementing various instructional approaches and

utilizing various combinations of resources. The "systems approach" allowed the Department to view all these needs in an interrelated manner.

A. STATE OF THE ART REVIEWS

1. Existing Testing Methodologies; Is the Educational System Succeeding?

The Department's review of existing testing methodologies revealed that existing testing methodologies are not adequate for addressing the first question outlined above--To what extent are New York State's students successfully mastering basic skills (e.g. reading, mathematics)? It has been generally accepted that standardized tests, which are the most widely used form of across-district testing in this State and throughout the country, are adequate measures to answer the question. This is not the case. Scores on such tests do not provide information concerning the outcomes of instruction because these tests are not designed to provide data concerning whether or not absolute standards of competence or the acquisition of particular skills have been attained.

Grade norms (i.e. scores based on group averages) from standardized tests are presently widely misinterpreted. It is widely believed by the public that norms from these tests define standards of competence for each grade, i.e. that they define an objectively determined level of performance on skills in a given subject area that all children in that grade should be able to reach. It is not generally understood that the "norming" process establishes averages requiring that 50 percent of all students be below grade level (i.e. average) when the test is standardized. The requirement that 50 percent of the students

fall below and 50 percent above is unavoidable because the methodological approach of standardized testing is to ascertain a student's performance in relationship to the performance of other individuals on a common scale based upon the range of performance of the norming group rather than to measure student performance in relation to a set of defined standards. In accordance with the purpose of ranking individuals in relation to one another, test items are purposely selected to spread out test scores. When these tests are given to students across the State or nation, they function predictably to spread students' total score out but do not show pupil performance in relation to a defined set of basic skills that constitute points on the continuum of learning to master a specific subject. If the concern is to measure the extent to which New York State students are mastering basic skills (e.g. reading) then logic would dictate that the evaluation system used should measure progress on each of those skills rather than on test questions that have been shown to have certain difficulty levels that demonstrate differences between individuals, and have been selected and scored or scaled on the basis of a national or state sample population of students.

Another problem associated with standardized tests is that they were not designed to measure adult standards. Some new testing methodologies in reading (and in some cases for other subject areas) are presently available to measure desired adult competencies: the National Assessment of Education Progress tests, Educational Testing Service Adult Reading Tasks, Harris Survey of "Survival" Requirements in Reading, etc. However, none of these

tests has been designed to measure progress toward adult reading competence or to help determine what progress seems to be appropriate in reaching those adult standards. Ideally, we would hope to have some specification of an ultimate standard or standards indicating what skill levels an individual needs to function as an adult. Furthermore, the Department has concluded that an appropriate test should measure at any time during a student's educational experience the extent to which he or she is progressing toward that standard or standards--that is, a test which indicates that, at some given time in the future, the student will be able to function according to the adult standards. Currently standardized tests do not allow educators to do either.

Experts set standards for a grade level, but there is no assurance that their judgment is correct in terms of adult standards. For example, if a twelfth-grade student obtains a score that places him at the twelfth-grade norm, there is no conclusive data concerning his ability to cope successfully with the reading tasks that he will meet in the adult world. All that can be inferred from the test score itself is that his reading performance, compared with the performance of others in his age group, is about average on a set of test items subjectively selected as appropriate for his educational level. The items are included in the test because a certain percent of the students at that age were able to master them when the test was piloted.

Another major discouraging aspect of the review has to do with the inability of standardized tests to adequately measure progress within a school year and to provide teachers, principals, etc. with adequate information to

make immediate and meaningful adjustments in educational instructional programs. In most cases, the items on standardized tests are too general or too few in number to serve as accurate measures of specific skill needs. Moreover, since the tasks used to measure general skills vary greatly from test to test, and even from level to level within a given series of tests, they are imprecise indicators of actual skill development and needs.

In addition to the general concerns expressed above, there are a series of other specific misunderstandings related to interpreting grade equivalent scores that must be understood by readers of this report.

Grade equivalent scores appear to be easy to understand and use. They have a "directness of meaning" that often causes them to be taken at "face value" without full consideration of the limitations imposed on them by procedures used in their derivation.

Some of the limitations are obvious. A beginning third-grade pupil who obtains a 6.0 grade equivalent score on a third-grade reading test does not read as well as the average sixth-grade pupil, nor should he be promoted to an average sixth-grade reading group or given average sixth-grade materials to read. His sixth-grade score (GE of 6.0) indicates that his reading ability is well above the average for third-grade pupils, probably because he has mastered the skills taught in his earlier grades in a more thorough fashion than most pupils in the nationwide sample on whom the norms are based. This grade equivalent score of 6.0, because it is scaled to reflect performance average of sixth-grade students, would have quite a different meaning,

of course, if the pupil had taken a sixth-grade reading test and, on that test, had obtained a GE score of 6.0. Thus, grade equivalent scores of identical numerical values do not indicate the same or identical levels of accomplishment and are not meaningful when considered in isolation from the content of the test and the group on which the norms are based.

Grade equivalent scores have other less obvious limitations. Descriptions of these limitations are included in most test and measurement books and articles, and therefore only a few pertinent ones are reviewed briefly in this report as reminders that grade equivalent scores are subject to misinterpretation and misuse.

1. Grade equivalent scores may not be comparable from one test to another in the same subject or in different subjects. Some tests distribute pupil scores over a wider range of grade equivalent scores than do other tests. For example, a pupil who obtains a score at the 60th percentile of sixth-grade pupils on each of two tests in the same subject (e.g. vocabulary) may obtain a grade equivalent score of 6.5 on one test and a 7.0 or 7.5 on the other. This is especially true of tests in different subjects, e.g. reading compared to mathematics, and is also true but to a more limited extent among different tests within the same subject.

2. Grade equivalent scores assume that growth is uniform through the school year. Because the scores for each month are graphed or interpolated from the one or possibly two points during the year when the tests are standardized, it is possible that a grade equivalent unit of one month (.1) obtained at one time during the year does not represent the same amount of achievement

as one attained at another time during the year. Pupils tested at a time other than when a test was standardized may have actual achievement slightly higher or lower than that indicated by their grade equivalent scores.

3. Grade equivalents lack uniform meaning from low to high grades. A grade equivalent score one year below grade level at grade three may represent a serious disability (may be at the 10th percentile) while a score one year below norm at grade eight may not be considered to represent a disability (might be at the 40th or 45th percentile, well within the average range).

4. Grade equivalent scores are misleading when used as measures of growth, except for pupils whose scores are at or near the average for the norm group. The expected gain of one year in grade equivalent score for one year of instruction is not realistic for pupils or groups of pupils whose achievement is outside the average range, and particularly for those at the extremes of an achievement range. For a school with below-average pupils a one-year gain in its median grade equivalent (for example from 2.0 for pupils when in the third grade to 3.0 for these same pupils when in the fourth grade) may well represent growth in achievement above normal expectation. Similarly, a one-year gain in median grade equivalent for a school with above-average pupils may represent growth in achievement below normal expectation.

There are additional deficiencies of a more technical nature which are not discussed here but it is clear that the scales presently available to measure the effectiveness of instruction are not able to serve the purposes

often attributed to them. These shortcomings obviously have confusing and distorting effects on "fair" reports of the effectiveness of the educational system, a particular educational program, or a component of an educational program.

The review of testing methodologies has led to a clearer identification of properties the Department desires for testing methodologies to be developed for the future. Progress in designing tests which will have these properties is discussed in Chapter II.

2. "State of the Art" Review - Availability of Disaggregated Information

In order to answer the second question outlined on page 2--What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?, a second major "state of the art" problem area has been examined below.

The problem stems from the lack of information presently available to help make decisions at various levels. In New York State it is estimated that there are 4,152,550 students enrolled in public and nonpublic schools. In addition, there are approximately 4,400 school buildings, 149,000 instruction rooms, 182,000 classroom teachers, and an unknown number of instructional programs being used by teachers. The problem is to determine efficiently what has transpired at the classroom level or between classrooms, so that successful instructional programs can be identified and replicated on a wide scale, and so that resources can be reallocated, when required, based on continuous evaluation. At all levels of the educational system--

local, state or federal--decision-makers in particular have been limited in their ability to identify successful instructional programs or to improve instructional planning, management and evaluation practices because of the lack of pertinent information.

The lack of capacity to make such information available has led to educational performance evaluations based on aggregate data of questionable relevance (e.g. gross expenditures by buildings, mean test scores by district and/or building, total numbers of students of different socioeconomic backgrounds in the district, etc.). The evaluations based solely on these kinds of data have severely confused or damaged public attitudes towards the educational system, since it has supposedly been shown that little seems to work in education and yet every parent realizes that their child is learning in schools. The state-of-the-art problem relative to standardized tests further aggravates the problem in that standardized tests:

> "often require skills or aptitudes that may be influenced to only a limited degree by experience in the classroom...since standardized tests ordinarily have a low degree of overlap with the actual objectives of instruction at any given time or place. Standardized tests, except for some individually administered diagnostic instruments, are designed to measure highly generalized skills. This inevitably means that scores on such tests have only a distant, even tenuous, relationship to the outcomes of the real-life instruction in a given situation. One cannot determine the effectiveness of educational programs or establish accountability in any meaningful sense if it is not clear that devices used to assess educational outcomes actually measure the objectives a given teacher, school, district or state is attempting to accomplish over a defined period of time." 1/

1/ UCLA - Center for the Study of Evaluation - Evaluation Comment, September 1971, Vol. 3, No. 1.

Attempts to combine less-than adequate outcome measures and aggregate information for evaluation purposes has led to some unfair conclusions regarding the effectiveness of educational programs. This unsatisfactory situation has been compounded by the fact that there often may be as much variation between buildings within a district or even within a building or grade as between any of the variables used for such analyses.

Given the state of the art in respect to the testing techniques described above and the availability of disaggregated information, it is apparent that questions related to the extent to which basic skills are being mastered and to the quality of education programs cannot be satisfactorily answered using techniques that are presently employed on a widespread basis. Consequently, a new evaluation system (i.e. the system that has been referred to in this report as "projected") has been designed and is in the process of being developed and implemented as described in the material below.

B. QUESTION #1: TO WHAT EXTENT ARE NEW YORK STATE'S ELEMENTARY AND SECONDARY STUDENTS SUCCESSFULLY MASTERING BASIC SKILLS?

1. Desired Purposes of Tests

The first task the Department undertook to determine a reliable answer to this question was to define what types of information would be required. Present efforts are focused on the development of new tests of reading ability (i.e. one basic skill) because of the importance of this skill. It is the Department's conclusion that there are three levels of testing required to measure this basic skill area (i.e. reading comprehension in the work done to date).

Question #1: To what extent are New York State's Elementary and Secondary students successfully mastering basic skills?

First, there is a need to determine how well a student performs on reading tasks that will be encountered in adult life. Presently there is a great deal of criticism of the educational system by employers to the effect that graduates of the secondary schools cannot read materials which are required to allow satisfactory performance as employees. In addition, to function in this society, it is necessary that individuals be able to comprehend income tax forms, driver's license applications, insurance policies, etc.

There is a need, then, for a testing methodology to determine in an efficient way (i.e. in a way that can sample adult reading tasks and test students at low cost) whether or not a student can deal with required adult tasks. Obviously, different levels of acceptable performance on such tests will be indicated depending on each student's individual aspirations. For example, a student desiring to become a lawyer will have to meet a higher standard of reading ability than one desiring to enter occupations where the standards might be lower.

A second level of testing is required in reading to measure progress toward adult standards over the entire school age range (i.e. beginning in primary grades, it is necessary to measure periodically to assess progress toward adult competence and to detect growth). The point is that, if the educational system waits until a student is close to high school graduation before testing adult competency, it is often too late to compensate for any ineffectiveness of the educational programs taken by the student, or to deal

Question #1: To what extent are New York State's Elementary and Secondary students successfully mastering basic skills?

with other problems. The new tests described in the Introduction to this report attempt to measure adult competencies, but have not attempted to incorporate this necessary additional component. The Department feels that it is a necessity to do so in its research and development activities in reading.

A necessary capability that is required for this second level of measurement is a technique that will allow reading materials to be scaled according to their difficulty. This will enable any materials including those read by adults and considered to be requisite to basic reading competence, to be scaled and located at a specific point. Likewise, reading materials typically encountered by students could also be scaled leading to the ability to equate student performance on the materials he uses to some point or points on a scale where the difficulty levels of adult materials are also located.

Using this technique students can be periodically measured during their school careers to determine their progress toward adult reading competence. Additionally, the scale will make it possible to determine the progress students and groups have made as a result of receiving instruction within a school year. Present measures such as standardized tests are not sensitive enough to determine reliable individual or group scores between fall and spring administrations of tests. This need is especially important in terms of making end-of-year evaluation, be it for categorical projects or regular instructional programs.

Question #1: To what extent are New York State's Elementary and Secondary students successfully mastering basic skills?

Furthermore, the scale will allow for comparisons between instructional treatments, including both variations in objective sequencing, teaching strategies and materials leading to a determination of which provide greatest growth. With these kinds of data resource decisions could be made based on data generated from comparative analysis across various levels of educational units.

The third level of testing requires a measurement device that, within a school year, can determine a student's success in mastering those subskills which are necessary to reach a given reading level by the end of a school year and identify those subprograms which are successful in realizing those subskills. Additionally, this level of testing is required to analyze the extent to which the subprogram as characterized by particular sequences of subskills is leading toward mastery of those skills for students with different characteristics. It is the Department's intent to follow student performance and subprogram objectives over time with a computerized information system (described under Question #2) so that successful patterns can be identified.

In summary, in the area of the reading comprehension, the Department has concluded that three levels of tests are required to fulfill educational decision-making needs. First, there is a need to determine how well a student performs on tasks that may be encountered in adult life. Second, the test(s) must be able to measure progress toward adult competence over grades and be sensitive enough to measure individual and program progress within a period

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

such as a school year. Third, there is a need for tests to verify that specific subskills result in a given level of competence at a specified point in time, and to verify the most effective sequences in which subskills can be acquired.

The three levels of testing identified above for reading comprehension are thought to be equally applicable to other subject areas.

It should be noted that neither the questions on or the description of past or planned work activities described below are directed at measuring the feelings, emotions, and attitudes that are acquired as a consequence of instruction and that are equally important as measuring basic skills. Local district personnel have expressed a great deal of concern that the Department is not doing more in this important area. The Department realizes the importance of developing adequate measures in this area but has concluded that, given the inadequacies of existing tests, the limited funds currently available for this total evaluation program, and the time it takes to develop measures of this type, that it is impractical to consider work activities in this area in the next few years.

2. Progress in Developing More Suitable Measurement Devices

The emphasis of the work completed or in progress regarding the first two levels of testing needs (i.e. setting standards for and measuring progress toward adult competency) has been in the priority area of reading.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

Present plans call for the development of similar tests in mathematics, if funds are available. Decisions related to other subject areas or to the measurement of attitudes, etc. have not yet been made.

a. The Development of a New Test Methodology

In the reading area, a one-year detailed research effort has been completed. The results of this effort were: 1) a determination of the properties desired in a new testing methodology (i.e. the reading effectiveness measure); 2) a description of the minimum number of broad development tasks which must be undertaken to develop the new reading effectiveness measure; 3) a review of existing approaches to the measurement of effectiveness in reading, both in terms of the properties desired and the technical problems associated with meeting the specified properties; and 4) a proposed work plan for developing the actual tests. The remainder of this section summarizes the results of this research effort under the four topic headings above.

1) Properties of the New Methodology: The properties identified in the report parallel the discussion above on the purposes of the needed testing methodology. The specific properties include the following:

a) A capability to quantitatively indicate the reading difficulty of the broad range of adult reading materials.

b) A capability to measure progress toward adult competence (i.e., a standard or standards). The test must be able to measure the progress of individuals and groups toward being competent adults.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

For example, in reading it should:

(1) Measure the ability to cope with societal reading requirements imposed by law, such as comprehending income tax forms or driver's license applications, and with other materials intended by government agencies for the protection and well-being of citizens.

(2) Measure the ability to read materials necessary to enter various vocations or professions.

(3) Measure the ability to read materials that enable individuals to function competently in their own behalf, such as advertisements, insurance policies, repair manuals, etc.

c) A capability to measure individual effectiveness. Since education is concerned with the development of individuals, the measure must yield reliable and valid individual scores.

d) A capability to measure system effectiveness. It must be possible to aggregate the scores of individuals (by grade, sex, ethnicity, etc.) to determine how well the educational system is performing for different target groups in different schools, buildings, school districts, regions, and statewide.

e) A capability to measure growth in reading ability. The measure should be able to detect small changes in reading ability, such as might be expected to occur in one year's time. Measurement of group growth is an essential requirement of the measure. Measurement of individual growth, if feasible, is highly desirable.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

f) A capability to measure reading ability over the entire school age range. Continuity of measurement, beginning in the primary grades, is necessary for measuring progress toward adult competence and for detecting growth. Therefore, the measure should be applicable over all or nearly all of the public school age range.

g) A capability to furnish meaningful scores. Scores on the measure should be readily and accurately understood by persons without technical knowledge of statistics or test construction procedures, such as parents, legislators, teachers, etc. Therefore, it must be possible to present scores in terms that are meaningful to such persons without sacrificing precision in reporting.

h) A capability to measure specific instructional objectives. Work efforts associated with the last property are described on pages 166-168.

2) Minimum Number of Development Tasks Required. The broad development tasks required to result in a new methodology to measure reading effectiveness are of two major types: First, there are legislative-political tasks. In this regard it is important to state that the actual setting of acceptable adult standard(s) will under any circumstances be a policy decision requiring some subjective judgments but data can be made available which can make that task of policy-makers more rational. Scientists can and should contribute sound, impartial technical work to describe the range of reading difficulty associated with adult reading materials and also project what reading demands will be in say, 15 years (i.e. when a student now entering school will graduate) but the judgment as to what standard or standards are to be required must be done by government.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

The second type of broad development task is categorized as a scientific-technical one. The specific scientific-technical tasks to be undertaken include: (1) identifying the various kinds of materials that adults are called upon to read (i.e., defining what is called the corpus); (2) scaling adult reading tasks - because of the wide range of reading materials in existence it is important to sample them in an efficient manner and determine the characteristics (e.g. number of words, sentence length, number of commas, etc.) of passages or frequency of use of words so that a formula can be developed to scale any material efficiently; and (3) carry out the steps necessary to test students (i.e. to develop tests) including such activities as selecting item formats, insuring that the test is reliable (i.e. that scores remain relatively stable upon retesting) and valid (i.e. does the test measure what it sets out to measure).

3) Review of Existing Approaches. The review of some of the newer tests being developed elsewhere indicate that none of these tests measures individual or group growth in reading achievement or progress toward adult reading competence. In addition, there are technical deficiencies in the tests when compared against the minimum requirements contained in the preceding paragraph.

The research report then discusses all of the technical alternatives associated with developing a suitable measure. It was concluded that readability and, to a lesser extent, vocabulary difficulty (i.e. incidence of rarely used versus frequently used words) are the properties to be measured.

In summary, it was concluded that existing tests will not yield the information required to answer the question--"To what extent are New York State's elementary students successfully mastering basic skills (i.e. reading in the work underway)?"

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

4) Proposed Work Plan. The following tasks are being undertaken in the immediate future to develop an appropriate testing methodology: (1) a criterion scale of readability will be developed from adult and school age materials using a technique technically called the cloze procedure (i.e. where students and adults select words that have been deleted from passages); (2) an improved formula will be developed, since the wide range of passages that must be scaled prohibit hand scoring of all passages (the formula will predict, based on the characteristics of any passage, what the readability is of that passage); (3) tests will be developed which test students' and adults' ability to read material of various difficulties. In addition, the computer will be used for sampling, for producing and scoring tests, and for reporting results.

To continue the development work on the "reading effectiveness measure" the Department has submitted a two-year proposal to the National Institute of Education (N.I.E.). Other outside funding will also be sought this spring if N.I.E. funding is not secured.

b. Other Uses for Data which will be Furnished by the New Test Methodology [1/]

There are significant uses of these data other than those for assessing pupil or program progress which are added incentives for developing the "reading effectiveness measure". For example, if students were not meeting standards as measured by the new tests, it would be important that State and local educational decision-makers have a way of identifying some of the factors contributing to the less-than-desired levels of effectiveness. Analyses such as those described below should contribute to an understanding of reasons why those standards are not being met, and indicate possible sources of corrective action. Perhaps all the factors contributing to a student's failure to attain such standards would not be identified, but significant information could be made available that would support constructive decision making.

1/ Design Concepts for a Measure of Effectiveness in Reading - A Feasibility Study, Riverside Research Institute, September 1973, pp. 179-192.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

1) Multivariate analyses of effectiveness in reading. Information concerning whether de facto grade standards (i.e. difficulty of materials used in a grade) were being met could be analyzed in conjunction with information concerning the pattern of progress toward adult standards over the grades in order to understand why any target group(s) or school system(s) were attaining a given level of effectiveness by the twelfth grade. Such a study would enable educational decision-makers to isolate problems and consider alternative solutions.

For example, let us assume that a problem definition study showed that all (black, white, and Spanish-speaking) middle class students attained the minimum adult standards by grade ten; all lower class (white, black, and Spanish-speaking) students did not attain minimum standards by grade twelve. Let us further assume that the de facto grade level standards increased by equal amounts each school year from grades one through five and then leveled off so that the rate of growth in expected reading skill was less for grades six through twelve than for grades one through five; furthermore, by tenth grade, de facto standards have reached the level of adult standards. Continuing with our hypothetical example, let us assume that all middle class students met or exceeded grade level standards in all grades, but that, starting in grade three, all lower class students did not attain grade level standards. It then follows that it would be reasonable to inquire into the feasibility of reducing, for lower class students, the rate by which the standards increased in grades one through five, increasing the rate in grades six through twelve, and setting the present de facto grade ten standard as the expected value in grade twelve. Such changes might reasonably be considered in order to better insure that lower class students attained adult standards by graduation.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

2) Analyses of instructional materials. The reading effectiveness measure could be used in analyses of instructional materials designed to locate possible problems in the readability and vocabulary learning demands placed on students. Once such problems were identified, responses designed to alleviate them, and hence to increase effectiveness, could be suggested to senior managers of the State Education Department and to local educational decision-makers.

The analyses that are illustrated below do not, by any means, exhaust the analyses of instructional materials that might be carried out in looking for sources of failure to meet adult standards. For example, the question of the consequences of alternative methods of teaching reading on attaining standards is not touched upon. Rather, the illustrations are limited to analyses that are related to the readability and word familiarity measures (i.e. subcomponents of the reading effectiveness measure [1]).

a) Disorder in instructional materials. The vocabulary of instructional materials is presently not set in any systematic sequential manner.

Evidence presented by Stauffer (1966) indicates that reading programs differ concerning which words are introduced in which grades. If there are major differences between the vocabulary in various sets of instructional materials, there should be difficulty in maintaining a logical sequence of instruction. Whenever a school changed books or whenever a student changed schools, students would be apt to encounter a great number of words not previously learned or, alternatively, be asked to learn words already mastered. The problem of encountering a great number of new words could be expected to be most serious for those students who are most dependent on the schools for what they learn, presumably educationally disadvantaged students.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

On the hypothesis that one knew what materials were used in all grades, the data would be available with which to determine the degree of similarity across reading program materials from different publishers with respect to the vocabulary and readability introduced in each of the grades. If a display of the overlap of readability and of vocabulary across programs confirms that, in fact, there is appreciable chaos (i.e., little overlap), corrective action would be indicated. A logical step to correct the problem would be for the State Education Department to direct publishers that there be (at least) a minimum specified amount of overlap in vocabulary and readability across reading programs.

b) The effect of the amount of learning expected per time unit (load) on reading development. There is evidence to suggest that the vocabulary load that is being placed on students is not being carefully planned or controlled. First, there is evidence suggesting that publishers do not coordinate vocabulary across curriculum areas. Stauffer's (1966) comparison of reading, arithmetic, health, and science books showed little overlap in each grade between the vocabulary words introduced in the books used in the different subject areas of the curriculum. For example, he found that while 2153 new words were introduced in seven reading series in the third grade, and 2150 new words were introduced in three arithmetic series in that grade, only 421 words were common to both lists. Moreover, many words appeared in textbooks in different subject areas which did not appear in any of the seven reading series at any grade level. Stauffer has estimated that even if a student somehow had the opportunity to learn the vocabulary of all seven reading series, he would learn only half the words he would encounter in his arithmetic books.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

If these data are correct, the effective reading load on pupils is considerably heavier than the requirements made by formal reading programs (and, for that matter, heavier than the requirements made by typical, college-level, foreign language courses), since students must also learn much new vocabulary in the subject areas. Furthermore, since new vocabulary is seldom identified as such in textbooks, other than basal readers, there is a good chance that teachers will not formally teach these new words, thereby leaving the learning of much critical vocabulary entirely up to the student.

There is also evidence suggesting that the reading load presently placed on students may not be evenly distributed over grades. Stauffer's (1966) data show a geometric increase in the number of new reading vocabulary words introduced per grade, in grades one through three.

The Department believes that loading rate should have important effects on reading achievement since a related variable, massed versus distributed practice, is a classic learning variable of considerable significance. However, the actual relations between loading rate and learning to read are unknown, and would need to be clarified in a program of research before the load currently found in instructional materials could be evaluated.

c) The effects of difficulty on learning. The readability formula and word frequency distributions could be used in analyses designed to determine whether the difficulty level of instructional materials being used by students might be responsible for their failure to attain adult standards of reading competence.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

It is well established that difficulty is a critical variable in learning. When materials are too easy for students, boredom results. When materials are too difficult, students may become frustrated, "tune out", and develop negative attitudes toward learning. In some cases, students may even develop negative attitudes toward themselves (e.g., "I am not capable of learning") or set unrealistically low achievement aspirations in order to reduce the impact of the failure experiences they are likely to have when asked to learn materials that are too difficult for them. In short, when materials are either too easy or too hard, students do not learn efficiently and the risk that educational processes will produce unwanted outcomes increases. It is for these reasons that controlling the difficulty of instructional materials has been a traditional concern of educators.

It is possible to examine these relationships once the reading effectiveness measure is available.

These possible analyses are discussed in this report to indicate that serious educational purposes beyond those of just finding out how students are progressing are feasible. It is one thing to determine progress of lack of progress but it is even more important to take corrective action. The Department feels linking this effort with the Instructional Evaluation System described on pages 154-166 will allow educators to do both.

c. Development of Basic Competency Examinations. One of the most difficult tasks the Department has faced in the implementation of the Evaluation Plan has been the effort to resolve the many issues that surround the development of basic competency examinations. Earlier the Department had indicated that during the 1973-74 year competency examinations would be developed in two areas: reading and mathematics. This plan included field testing of these two measures by June of 1974 and making the examinations available for general use in the schools during 1975.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

Originally the Department indicated that competency examinations would be provided to local school districts for use with all students following the eighth grade. It was thought that the examinations could be taken at any time after the eighth grade and as often as need be to assure that students reached a certain basic competency level by the time they are graduated from high school. The Department's plan was to have competency examinations not only in reading and mathematics but, in the future, in the areas of social studies and science as well. The latter two areas were not planned to be developed during the current year.

Perhaps the most significant concern the Department has had to deal with in this task is the matter of agreement upon the definition of basic competency as it would be related to students in the schools of the State. Two of the basic questions that had to be addressed were (1) would the measure deal with the progress the student makes toward adult competency, or (2) would the test deal solely with competency level as it relates to curriculum through the eighth grade. The development of a measure that will be challenging to all students at this level is extremely difficult. Therefore, the Department is weighing the advisability of providing exemptions from competency examinations for certain youngsters, i.e., mentally retarded, students who have scored above an established statewide reference point on the pupil evaluation program test, or students that may be pursuing Regents level courses.

The Department has involved the field in this discussion to a considerable extent. The development of basic competency examinations has proven to be controversial with a considerable segment of the field. Many cautions have been expressed which should be very carefully reviewed and considered before any basic competency examination is mandated for general use in all schools of the state. The Department will continue its efforts to develop competency measures taking into account the concerns and cautions that have been expressed by the field on this issue. In the meantime, interim measures have been under consideration for possible use.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

There were three alternatives considered as short-term measures for the basic competency examination in reading and mathematics. First, in the reading area, it is possible to utilize the Educational Testing Service (E. T. S.) measures of minimum adult competence or other commercial examinations. In the work of E. T. S. a national sample of adults were asked what they read in an average week. Various studies and judgments were then made as to what reading activities were most important, and related questions were prepared. Although this measure does not incorporate all of the properties suggested for the reading effectiveness measure nor does it follow all the steps that are necessary to develop a new measure, it is a reasonable short-term plan that will be considered.

A second alternative would be to use an experimental version of the "reading effectiveness measure". With this alternative, existing materials used in high school would be accepted as de facto standards of "basic competency". A sample of textbooks could be selected and its "readability" calculated with all existing formulas (the Department realizes there are deficiencies with these formulas but they can be used to scale materials used in secondary schools in New York State). The use of "fairer" tests could also be encouraged by holding the difficulty of test items constant for a selected passage with a given "readability".

The third alternative to satisfy the need for an "interim" basic competency examination would be to have experts select instructional criterion-referenced objectives and test items from the Department's bank of behavioral objectives and test items in reading or from similar commercial banks. Obviously, this would be much more subjective than the ultimate measure, but it is an alternative. This might be considered preferable to utilizing a standardized test, since the purpose of these tests is to compare students, not to determine whether explicit standards (here called basic competencies) are achieved.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

d. The Implementation of the Instructional Evaluation System. Up to this point, the third level of testing required has not been discussed other than to indicate that there is a specific need that subskills be measured on a continuous basis. One major component of the Instructional Evaluation System (described in this Chapter under the second question "What types of instructional programs are more effective for students of various characteristics?") is the development of behavioral instructional objectives (i.e. a statement of what a student should be able to perform after receiving instruction on subskills) and the development of criterion-referenced test items for measuring whether or not students and/or programs have been successful in achieving behavioral objectives.

The Instructional Evaluation System also provides test scoring services and feedback information to educational decision-makers through a computer support capability. In the context of the question being addressed here ("To what extent are New York State's elementary and secondary students successfully mastering basic skills?"), it is only necessary to indicate that, at the instructional level, basic skill is defined in terms of subskills which informally have been identified over the years as appropriate to reaching particular levels of reading comprehension. The point is that with the I.E.S. the appropriateness of particular subskills in contributing to reading comprehension will be determined and that criterion-referenced tests will be used to measure, as an instructional program is being implemented, whether particular subskills are being learned are being developed. [1]

1/ The use of instructional criterion-referenced objective tests to measure performance on objectives is rapidly increasing in this state and elsewhere, but it does take time to implement a system of instructional criterion-referenced objective tests on a large scale. The Department's progress in implementing its two approaches of I.E.S.-- the Instructional Support System (I.S.S.) and the System for Pupil and Program Evaluation and Development (SPPED)--will be described in the next section under the second question.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

e. Pilot Study of Instructional Criterion-Referenced Objective Tests. Throughout this discussion of testing need, the Department has indicated that standardized tests are not as useful as criterion-referenced tests, especially at the instructional program level. To determine the feasibility of supporting criterion-referenced testing on a broad scale in all the schools in New York State, a pilot study has been undertaken during this school year (1973-74) in ten school districts with different needs for reading programs, grades 4-6.

Since the assumption is that the criterion-referenced approach will be a more effective procedure than the standardized test approach for assessing the effectiveness of instructional programs at the school district, building, and classroom level, this study is examining in detail how the relative sensitivity of the two types of testing is affecting differences between schools in how reading is taught, time spent on reading, the materials used for instruction, and in the number and types of additional personnel involved in reading instruction. It is the Department's expectation that the more specific and relevant data available from criterion-referenced tests should enable district personnel to manage the instructional program more effectively and thus facilitate student achievement.

A broad range of district types are participating in the study. The names of these districts and their contributions to the total study sample are shown in Table 42. Also presented are the number of levels and the criterion-referenced test forms being developed for each participating district. Generally, a span of three grade levels is required because the range of capabilities of students in any school district grade varies.

Each level of testing is typically composed of 5 different criterion-referenced test forms tailored to the stated reading program in a district. This pattern of levels and forms for the criterion-referenced reading tests is in marked contrast to the use of a single level and form for three or more grade levels in the administration of a standardized reading test.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

Table 42

Grade Level Sample for Participating Districts

District	Grade Levels	Sample Size	Test Levels	Test Forms	Starting Date (73-74)
Brentwood[a]	4-5-6	1000	5	20	Sept
District 28[a]	4-5-6	400	6	30	Feb
Greece	4-5-6	1500	3	15	Jan
Jamesville DeWitt	4-5-6	1000	6	30	Jan
Plattsburgh	4-5	400	4-5	20	Feb
Rush Henrietta	4-5-6	500	6	30	Feb
Syracuse[a]	4-5-6	2500	6-7	30	Feb
Utica[a]	4-5-6	1200	5-6	30	Feb
West Seneca	5-6	800	2	15	Sept
Yonkers[a]	4-5-6	700	5-6	30	Jan
Totals		10,000		250	

[a] These districts have a common standardized achievement test (CAT) and will be combined for some analyses in the study.

Figure 1 shows the major activities and schedule of the study. Districts selected objectives in the areas of reading comprehension and vocabulary using the newly available comprehensive New York State bank of reading objectives. The objectives in each of these areas were "leveled" into the range of reading levels found among students in grades 4-5-6 in a district. Lists of reading materials and word lists were then associated with the reading objectives in each level. These and other data constituted the specifications and raw material used for generating some 750 to 1200 test items for the criterion-referenced test forms for each district. Typically, each district will have 5-6 test forms at each level, with 30-40 test items (half in the area of vocabulary and half in the area of comprehension) on a form. There are, then, some 150-250 test items available for each level of testing.

Question #1 : To what extent are New York State's elementary and secondary students successfully mastering basic skills?

Figure 1

Schedule of Study Activities (1973-74)

Activity	S	O	N	D	J	F	M	A	M	J	J	A
Selection of Objectives	x	x	x									
Prepare CRT Test Specifications				x								
Develop Test Forms				x	x							
Administer CRT's					x	x	x	x	x	x[b]		
Administer Standardized Tests		x[a]			x				x[c]			
Return CRT Reports						x	x	x	x	x		
Analyze Data										x	x	
Prepare Report												x

[a] PEP test collect from the sixth grade sample.

[b] Administration of adult reading competency measure developed by ETS to 6th graders only.

[c] A variety of different standardized tests scores will be available, some with only post test scores, whereas others will have as many as three sets of scores for the school year.

The administration of the criterion-referenced tests shown in Figure 1 begins in January of 1974. Students in a given classroom will receive the test forms on the appropriate levels in random order on a monthly basis until they have received all five forms in a level. All levels of tests developed for a district may be used in a classroom contrasting sharply with the typical administration of a single level of a standardized test in all classrooms in a grade.

The remaining elements of the schedule shown in Figure 1 indicate the return of classroom and district level reports, on a monthly basis, using the data processing system called CAM (Comprehensive Achievement Monitoring) described under the Instructional Evaluation System section of this report. Also shown are the analytical and reporting stages of the project.

Question #1: To what extent are New York State's elementary and secondary students successfully mastering basic skills?

The study has already yielded some important data on the feasibility of using criterion-referenced testing on a statewide basis. That is, it is obvious at this point that the participating districts have largely selected the same instructional objectives in the areas of vocabulary and comprehension. Where the districts generally differ most is in the "content" associated with the objectives and in the number of testing levels applicable to the children in a given classroom or building. Urban districts tend to use the full range of test levels that can be made available and also add one more lower level of testing compared with suburban or advantaged districts.

This experience with creating several levels of criterion-referenced tests for a district suggests that some 15-20 levels of testing could be created to cover the full range of reading levels generally demonstrable in grades K-6. An acceptable plan for evaluating programs in districts would allow for assigning each individual student to the appropriate developmental test level, regardless of grade level. A district may then assess the effectiveness of its programs in relation to the progress of students from level to level.

The technology created to support the development of criterion-referenced tests in reading has so far proved basically adequate for the task. The bank of reading objectives is currently being computerized and made available statewide. Test items for the development of criterion-referenced tests in reading will also be indexed by computer to support economical test development on a much larger scale in the future. As experience in criterion-referenced test development continues to be gained in the study, the ability to support this mode of testing through increased use of technology appears more and more feasible.

In 1974-75, this study will be expanded in the same districts to cover grades 1-6 in reading. The statewide objective and test item bank will be expanded. In 1975-76, packaged materials for use in all districts will be completed.

In summary, the question To what extent are New York State's students successfully mastering basic skills? can be answered in different ways depending on how "basic skills" are defined. In this "projected" program three levels are

seen as being required. Progress toward developing suitable measures is progressing well on "interim" activities planned or implemented to best answer the question in the interim period. A more complete review of the "interim" program was discussed in Chapter II.

C. QUESTION #2: WHAT TYPES OF INSTRUCTIONAL PROGRAMS (OR PROJECTS IN THE CASE OF CATEGORICAL AID PROGRAMS) ARE MOST EFFECTIVE FOR STUDENTS OF DIFFERENT CHARACTERISTICS?

The Department's efforts described in Section II A have a direct relationship to this question. Education must have adequate measures of basic skills before comparing alternative approaches (e.g., instructional programs or other "input" quantities) to try to respond to this question. In addition to resolving the testing problem, the Department is convinced that improved methods are required to obtain explicit information about programs in order to be able to identify which instructional programs work better than others.

At present, as indicated in the introduction, it is not feasible to trace comprehensively what approaches are being used in various districts, buildings and classrooms throughout New York State. The practical problems inhibiting the collection of meaningful information have forced educational evaluators to use a variety of aggregated information. Even in the case of categorical projects funded under the State's Urban Education Act and E.S.E.A. Title I, where there is considerable more documented information on what is supposed to happen if projects are funded, seldom is there explicit information on what instructional activities will be used. In addition, there generally is no practical method of tracing whether or not project elements were implemented as proposed and whether or not some or none of the parts of the instructional

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

program were successful for students of different characteristics. Improving this situation is equally as important as developing more useful testing methodologies.

Eliminating these practical problems is a tremendous undertaking given the number of school districts, buildings, subjects taught, classrooms, and students in the State. The Department now feels that this can be done by gradually implementing the Instructional Evaluation System (I.E.S.).

The I.E.S. is now being implemented with two different computer software systems--one called "C.A.M." (Comprehensive Achievement Monitoring) and the other called "I.S.S." (Instructional Support System). A decision was made some three or four years ago to develop these two systems because, even though the I.E.S. concepts were generally accepted as being needed, no one really knew how best to operationalize them, in particular in terms of the computer software system requirements.

Surprisingly, both systems, although developed and pilot tested independently, are very similar. With the growing acceptance of both approaches in an increasing number of districts, it is clear that the most practical course of action is to use both because the resources are hard pressed in both cases to serve the whole State in the next few years. Generally the C.A.M. system is being implemented in upstate and Long Island districts and the I.S.S. in New York City. The I.S.S., however, was piloted in an upstate district.

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

It should be noted that the Department believes that the implementation of the Instructional Evaluation System is also required to improve instructional planning, management and evaluation practices in local school districts--i.e. to answer question #3. The description which follows of the concepts of the Instructional Evaluation System should make it clear that this reasoning is sound.

1. The Instructional Evaluation System - Overview of Purposes

The Department believes that standards relating to the following aspects of the instructional environment need to be established and met if Question #2 is to be answered. As indicated below, all of these standards are not now being met.

1) quality of instructional objectives;

2) quality of instructional program designs (i.e., well-designed and documented instructional programs);

3) appropriate test instruments selected to measure performance; and

4) utilization of information from the instructional program for within-school-year and between-school-year decision-making.

a. Instructional objectives

Objectives must be well formulated and must be related to the measurement tools selected and vice versa. Over the past several years, program objectives have been improved in categorical project applications and also in regular school programs. Three additional improvements, however, are required in many districts.

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

1) Objectives continue to be stated for the whole program (end-of-year program objectives) rather than for portions of the program (e.g., two-week groups of learning activities that are generally called modules). This present reliance on end-of-program objectives precludes analysis of parts of programs. For example, a program might not be too impressive in terms of the final results obtained, but many components of the program might well be effective. Documentation of module objectives is also needed if continuous program adjustments are to be made while the program is in progress rather than when the end-of-year evaluation has been completed.

2) Too many project objectives are stated in terms of results expected from standardized tests (e.g., the participants will advance from the 10th to the 25th percentile on the test). These types of objectives are popular because standardized tests are readily available. An analysis of some instructional programs indicates that these tests are frequently unrelated, to varying degrees, to the subject material being taught. Therefore, this combination of objective type and the test instrument employed has frequently been found to be inappropriate for the evaluation of instructional programs. (This barrier was discussed in the Introduction.)

The Department is assisting local districts in developing district-made mastery tests which are closely related to the objectives of each instructional program. Results of standardized tests are also being recorded to analyze the relationships between the two testing methods until replacement tests (i.e., tests like the Reading Effectiveness Measure) are developed.

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

3) The existing practice is to document objectives for categorical programs only, and not for regular ongoing programs. The impact of a categorical program cannot be properly determined unless comparable information about regular, ongoing programs given in the same school district is available. In those districts where the full IES is being implemented, the Department is assisting districts in developing module objectives related to instructional activities for all programs. These objectives are not stated in terms of expected standardized test results, but instead are stated in terms of the behaviors to be developed as a result of instruction which are in turn assessed by criterion-referenced mastery tests.

In districts where the full IES is being implemented, the Department is supplying technical assistance to insure that objectives are properly defined. In these districts, a computerized support system is provided to efficiently store the objectives for management and evaluative purposes.

b. Instructional Program Designs

The second standard is the quality of program design (i.e., having a well-designed and documented instructional program). Progress has been made in having districts define their programs, but learning activities grouped into modules must be more fully documented if the local districts or the State are to know to what extent particular modules were effective in teaching different groups of students. Currently, most school districts do not have the necessary computerized capability for documenting instructional programs efficiently, either for regular or categorical programs.

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

In districts where the IES is being implemented, the computerized support system is available for storing program information for management and evaluative purposes. The Department is rendering technical assistance to assist pilot districts in developing modules and objectives (i.e., learning activities grouped according to specific objectives and covering roughly two-week periods of instruction) for the entire educational program (e.g. mathematics).

c. Utilizing Information for Decision-making (i.e. feedback)

Currently, neither teachers nor principals receive timely information that would permit adjustments to be made within and between classrooms. This situation arises because it is physically impossible to manually process all the necessary information on instructional objectives, program modules, module tests, end-of-semester and end-of-year tests for all students and all subject areas. The computerized support system can make all this information readily available. With this capability, principals would find it possible to assume a larger role in instructional management and evaluation.

In those districts in which the IES is being implemented, the Department is rendering technical assistance to teachers and principals in assisting them in utilizing information for decision-making. With the IES, it is possible to make program adjustments before students have progressed to a point where intervention becomes very difficult.

With the implementation of the I.E. System in more districts, the Education Department will receive much more timely information, and reporting to the Governor and the Legislature can be improved because of the capability provided by the computerized support system.

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

The implementation of the I.E.S. as described briefly above obviously will allow both local and State decision-makers to answer question #2--What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

Since the Department would have access to program and module information on a continuous basis in those districts which have implemented the full IES, the capability to determine the comparative effectiveness of modules and/or of total programs for the disadvantaged would exist. Information on successful programs can be disseminated, and local districts can be warned if some approach had failed elsewhere.

d. Computerized Support System

Because the implementation of the I.E. System is strongly dependent on the computer support system mentioned above, a summary of the system which has been developed and implemented is presented next.

The New York State Education Department has acquired rights to a flexible software system which has unique features (I.S.S.). The system will allow local education agencies (LEA's) to build pupil and instructional program files as part of ongoing administrative and instructional activities. With this system, districts can record comprehensive information on their instructional programs as well as information needed for administrative and reporting purposes (e.g., attendance, scheduling, transcripts, and reports to the community, parents, students, the State, and the Federal government). Presently, to fulfill reporting requirements, most districts must either manually lift data from many paper filing systems or ask teachers to provide the information. In either case the burden is much too great, especially for teachers. Furthermore, such "survey" procedures to recover manually filed data are inefficient, and may result in inaccurate reporting.

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

The unique features of the computer support system which the Department is implementing for IES include:

- capability to handle extensive amounts of data on students, and programs at a reasonable cost;
- capability to add or delete elements of information at will without redesigning the system (thus users are not locked into fixed data files);
- capability to produce a variety of reports (users can change the format and content of reports and "massage" data from many files at will);
- ease of use by laymen (a layman's guide is provided to enable users to address the computers without special training);
- capability to enter data easily as part of ongoing operations.

e. Summary of the I.E.S. Components

The I.E.S., in summary, consists of two principal components. First is the computerized information-handling component which scores tests, keeps records on pupil and programs and generates reports for students, parents, teachers and educational decision-makers. The State can gather data from local school districts on successful educational programs. It is important to emphasize the State will not receive information on specific students. It is only interested in programmatic information.

The second component of the I.E. System is a set of procedures at the local level for:

1) systematically developing instructional program objectives that can be operationalized across classrooms;
2) systematically developing or assembling instructional activities and mastery test items linked to instructional program objectives;

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

3) systematically measuring student performance with mastery tests, and interpreting test scores using predetermined performance criteria;

4) systematically reallocating instructional resources (e.g., materials, teacher time) to improve student performance; and,

5) systematically recording instructional activities, objectives, and test scores to provide a basis for upgrading instructional program content and evaluating instructional performance over time.

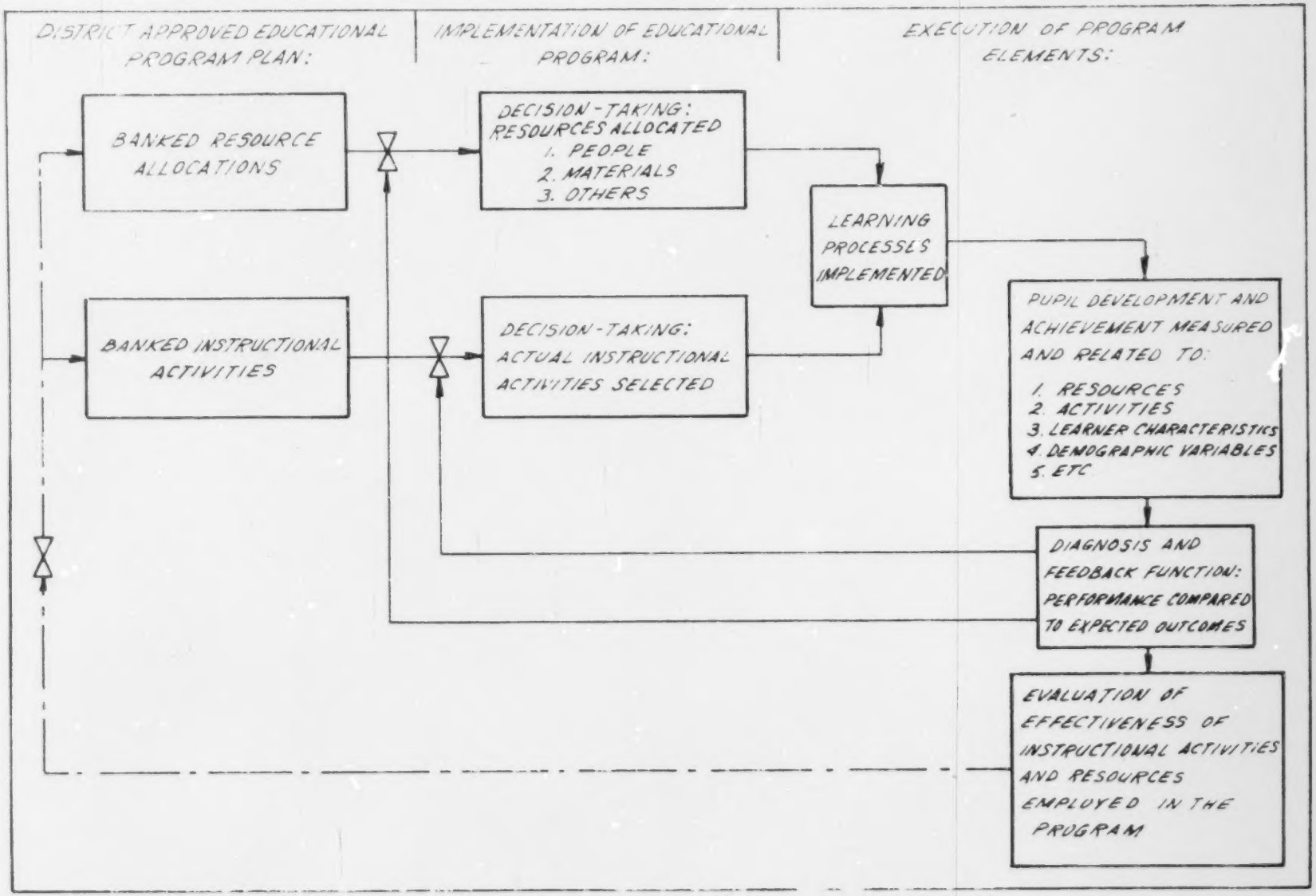
These two components function together to provide feedback information to teachers and other educational decision-makers so that student performance can be improved by:

- detecting discrepancies between planned and actual performance by students
- taking immediate corrective action with respect to instructional resources being allocated to and instructional activities being implemented in the classroom
- taking long-term corrective action with respect to the supply of instructional resources and activities made available to classroom teachers.

The feedback is represented on the accompanying diagram. One starts out with decisions on resources such as who will teach whom, what rooms will be used, what instructional activities are to be used, etc. and then traces general resources and instructional activities as they are used. After classroom activities are implemented, tests are given and continuous information is made available for a number of purposes such as trying alternative learning activities if the original one fails with a particular type of student, documenting successful approaches, utilizing reading specialist's time more efficiently and effectively and for many other decision-making purposes. In addition, reports to parents and the community can be made. The feedback loop provides LEA's with information that has not been available in the past.

Figure 2

THE INSTRUCTIONAL EVALUATION SYSTEM (IES) AS A CLOSED LOOP FEEDBACK SYSTEM


DISTRICT APPROVED EDUCATIONAL PROGRAM PLAN:
IMPLEMENTATION OF EDUCATIONAL PROGRAM:
EXECUTION OF PROGRAM ELEMENTS:
BANKED RESOURCE ALLOCATIONS
BANKED INSTRUCTIONAL ACTIVITIES
DECISION-TAKING: RESOURCES ALLOCATED
1. PEOPLE
2. MATERIALS
3. OTHERS
DECISION-TAKING: ACTUAL INSTRUCTIONAL ACTIVITIES SELECTED
LEARNING PROCESSES IMPLEMENTED
PUPIL DEVELOPMENT AND ACHIEVEMENT MEASURED AND RELATED TO:
1. RESOURCES
2. ACTIVITIES
3. LEARNER CHARACTERISTICS
4. DEMOGRAPHIC VARIABLES
5. ETC
DIAGNOSIS AND FEEDBACK FUNCTION: PERFORMANCE COMPARED TO EXPECTED OUTCOMES
EVALUATION OF EFFECTIVENESS OF INSTRUCTIONAL ACTIVITIES AND RESOURCES EMPLOYED IN THE PROGRAM

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

f. Evaluation of I. E. S. -Based Instructional Programs

The question of whether the implementation of an I. E. System in a school district makes a difference in terms of student performance is difficult to assess since standardized tests are the only measurement device currently available to measure success. This report has already indicated that its sensitivity to actual classroom instruction is questionable. Even with this constraint, a pilot district in upstate New York has used standardized test results as a measure of success because it felt (and the Department concurred) that an evaluation of the system was required. This evaluation is not completed but it is presented here to indicate some measure of the effect of implementing the I. S. System. The complete evaluation will be available in approximately six months.

Data available at this time include the results of the California Achievement Tests in reading and mathematics administered across grades, to different students in each grade, before and after the installation of ISS, e. g. grade 2 before ISS and grade 2 after ISS. The results of these tests are shown in Tables 43, 44, and 45. While the data presented have not been subjected to statistical analysis their implications are especially significant when related to the manner and intensity with which ISS was implemented.

Throughout 1972-73, implementation was stressed most heavily in the primary grades, next in the intermediate grades and lastly, in the high school. It is interesting to note that the data shown in Tables 43 and 44 reflect these activities in that the trend in gains or losses are consistently positive in the primary grades but appear mixed in the intermediate school grades. The significance of this finding should not go unattended.

Table 43

Mean Reading Achievement Before and After ISS in Guilderland Central Schools

Test	Grade	Before ISS Means: Expanded Score	Before ISS Means: Converted Grade Pt.	After ISS Means: Expanded Score	After ISS Means: Converted Grade Pt.	Trend + Gain - Loss
			Primary Grades			
Vocabulary	2	367.8	3.3	373.1	3.6	+
Comprehension	2	382.0	3.3	390.4	3.5	+
Total	2	360.7	3.3	368.4	3.6	+
Vocabulary	3	404.2	4.6	409.5	4.6	+
Comprehension	3	430.8	4.5	445.1	5.2	+
Total	3	408.8	4.6	421.8	5.1	+
Vocabulary	4	426.2	5.3	432.3	5.5	+
Comprehension	4	468.5	5.5	467.6	5.5	-
Total	4	437.0	5.5	440.1	5.6	+
Vocabulary	5	463.0	6.6	469.6	6.8	+
Comprehension	5	496.2	7.1	506.6	7.3	+
Total	5	471.9	6.8	480.7	7.0	+
			Intermediate School			
Vocabulary	6	498.1	7.5	501.6	7.7	+
Comprehension	6	512.3	7.6	514.6	7.6	+
Total	6	500.0	7.6	503.5	7.7	+
Vocabulary	7	536.6	8.6	533.1	8.6	-
Comprehension	7	560.2	9.5	550.7	9.1	-
Total	7	546.2	9.0	538.7	8.8	-
Vocabulary	8	577.3	10.1	579.2	10.1	+
Comprehension	8	585.0	10.3	597.8	10.6	+
Total	8	580.0	10.1	589.5	10.4	+

Table 44

Mean Mathematics Achievement in Guilderland Central Schools

Test	Grade	Before ISS Means: Expanded Score	Before ISS Means: Converted Grade Pt.	After ISS Means: Expanded Score	After ISS Means: Converted Grade Pt.	Trend + Gain - Loss
			Primary Grades			
Computation	2	310.2	3.2	313.5	3.3	+
Concepts	2	358.9	3.4	363.6	3.6	+
Total	2	315.1	3.3	319.4	3.4	+
Computation	3	361.7	4.5	363.2	4.5	+
Concepts	3	402.8	4.9	412.1	5.2	+
Total	3	368.7	4.7	373.6	4.8	+
Computation	4	397.5	5.2	400.9	5.4	+
Concepts	4	429.5	5.6	434.2	5.9	+
Total	4	398.2	5.3	402.2	5.4	+
Computation	5	434.4	6.0	440.2	6.1	+
Concepts	5	463.2	6.9	470.6	7.2	+
Total	5	434.4	6.2	441.2	6.4	+
			Intermediate School			
Computation	6	463.0	6.6	484.4	7.3	+
Concepts	6	485.6	7.4	488.4	7.6	+
Total	6	467.7	7.1	480.7	7.3	+
Computation	7	515.3	8.1	538.3	8.5	+
Concepts	7	545.4	9.0	551.4	9.3	+
Total	7	526.0	8.5	541.0	8.9	+
Computation	8	562.9	9.4	576.4	9.7	+
Concepts	8	601.9	10.7	591.1	10.4	-
Total	8	580.0	10	581.6	10.0	+

Table 45

Percent of Students Achieving Below Aptitude in Reading and Mathematics[a]

Academic Year	Reading	Math
1971-1972 Before ISS	9.0%	9.0%
1972-1973 After ISS	6.8%	8.1%

[a] "Below aptitude" is defined as a meaningful negative difference between a student's achievement score and his or her expected achievement score based upon IQ.

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

The installation of the ISS in an upstate district was unique in many respects, including the very important fact that the system is totally dependent on classroom teachers for the development of materials and definition of instructional strategies. In contrast, other attempts to define and implement such systems have relied heavily on predesigned software and control by outside consultants. Experience has shown that intimate teacher involvement is a prerequisite for success in such endeavors and these data support that conclusion. They also support the contention that unless one is prepared to do something intensively it shouldn't be done at all.

Table 45 shows the Percent of Students Achieving Below Aptitude in Reading and Mathematics before ISS and after ISS, across all grades. It can be seen that, before the introduction of ISS, 9.0 percent of the students could be classified as "below aptitude" in reading; after ISS, only 6.8 percent were counted in that category. In mathematics, the figures were 9.0 percent before and 8.1 percent after. In both instances we again see a trend toward positive improvement.

While the data presented are, admittedly, preliminary, it does seem reasonable to conclude that the I.S. System shows definite signs of success.

2. Progress in Implementing the Instructional Evaluation System

As indicated in the evaluation just discussed, one upstate school district has over the past three years implemented the I.S. System. This project was supported by federal funds. This funding will terminate this year but the school board committed itself some three years ago to continue it if results were positive. It appears that it will be continued.

In the present year the highest priority is to begin working with New York City community districts. One community district has already committed itself

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

to implementing the I.S.S. in all of its thirteen elementary schools for mathematics in grades 1-6. This decision was made after several trips to the upstate pilot by the superintendent, central staff, principals, supervisors and program managers. In New York City meetings have been held with the community school boards, union representatives, teachers, parent groups and an advisory group established specifically for I.S.S.

This district realizes that a tremendous amount of work is required at the outset to organize the curriculum, develop the learning activities into modules (approximately two weeks of instruction), develop test items, and set acceptable levels of performance but feels that the payoff in terms of iterating sound programs for students once implemented by having feedback information is worth the effort. (Appendix A provides a description of activities that make I.E.S. an operational system.)

The Department, through consultants, will provide technical assistance to the participating districts in carrying out these activities and will adjust the computer support system to improve its efficiency.

Additional community districts have been contacted and given detailed explanations regarding the I.S.S. It appears that one additional district will become involved in the next few months.

In upstate and Long Island districts, the CAM (Comprehensive Achievement Monitoring) software package is already being used in many districts.

Question #2: What types of instructional programs (or projects in the case of categorical aid programs) are most effective for students of different characteristics?

Efforts are underway to expand the coverage through B.O.C.E.S. personnel. In addition, improvements have recently been made to the computer software package to increase its ability to trace learning activities, handle biweekly mastery tests and to generate additional reports to local personnel.

In summary, the Department is making substantial progress in providing local districts with an ability to determine what instructional programs are more successful than others for students of different characteristics. The availability of such information at the State level will clearly enhance our capability to disseminate information to other school districts and to the Governor, Legislature and the Board of Regents.

D. QUESTION #3: TO WHAT EXTENT ARE THE BEST INSTRUCTIONAL PLANNING, MANAGEMENT AND EVALUATION PRACTICES BEING USED AT DIFFERENT LEVELS OF THE EDUCATIONAL SYSTEM?

1. Instructional Evaluation System.

The Department has two major efforts underway to address this question. The first effort is connected with the implementation of the Instructional Evaluation System. As the Instructional Evaluation System is implemented in school districts, new methods of involving teachers and other district personnel in structuring and developing subject-matter curricula (i.e. instructional planning) are being explored. In addition, as districts are provided with previously unavailable information regarding the success of various instructional approaches for children with different characteristics,

Question #3: To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?

new roles and responsibilities are often required of teaching and supervisory personnel (i.e. instructional management and evaluation). Various techniques are being used. In essence, the Department's plan is to iterate toward improved instructional planning, management and evaluation practices as needs become known.

2. Categorical Funding Study

The second effort is one of systematically examining instructional planning, management and evaluation practices in a sample of school districts presently receiving funds for compensatory education programs (i.e. ESEA Title I and Urban Education). This study is called the Categorical Funding Study. Some preliminary findings of one part of that study are described below. The final report will be made in the fall of 1974. The Department believes that the recommendations in the final report will be generally applicable to all educational practices, not just to those connected with compensatory efforts.

a. Purposes of the Study

In order to more fully understand the preliminary findings reported below, all work activities underway in the Categorical Funding Study, to be included in the fall report to the Governor and Legislature, must be outlined. The goal of the Categorical Fundirg Study is to recommend feasible, cost-effective means for the State Education Department to improve the planning, management, and evaluation of compensatory programs for the educationally disadvantaged. This goal arises from analyses suggesting

Question #3: To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?

that relatively high quality instructional planning, management, and evaluation processes will assist in achieving the desired outcomes of compensatory education. There is a realization, however, that the improvement of these practices is not the only change needed to achieve better outcomes for students (other elements include the people involved, the educational approaches used and the like).

The rationale for the study was threefold:

First, categorically funded compensatory education programs have been attacked as costly, ill-managed, and ineffectual, leading to proposals for their elimination or radical restructuring. Because many proposals for changing aid programs assume a high quality of planning, management and evaluation skills and practices at the state and local levels the Department has concluded that this assumption should be tested in order to improve decisions in the future.

Second, compensatory education is funded by a variety of grant mechanisms. Because much of the Department's programming involves grant programs, systematic study of the relative merits of alternative grant structures and grant management techniques would undergird many broad-ranging policy issues. This is especially important at this time since at the State level there are under consideration proposals for applying a weighting factor to the State aid formula for educationally disadvantaged students.

Third, improved understanding of existing local school district planning, management and evaluation processes is prerequisite to the design of comprehensive, improved systems for planning. management and evaluation. Because compensatory programs require more documentation than many other programs, study of them permits preliminary but systematic appraisal of local processes.

The Categorical Funding Study consists of six major, related components. Below is a brief description of the objectives, procedures, status, and expected outcomes of each component:

Question #3: To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?

1) Study of the Situation. The purpose of this component of the study was to define the broad problems of compensatory education with sufficient precision to carry out the other components. This involved a series of library and original research efforts that culminated in the design of the balance of the Project. The results of these efforts are currently being documented. The direct benefits of this component will be review of the relevant literature and a clearer understanding of the problems in compensatory education.

2) Functional Specifications for the Desired System. This component is intended to define a coherent set of specifications of the processes and structures that will contribute to the effectiveness of compensatory education. A synthesis of contemporary thinking regarding planning, management, evaluation, organization, and systems engineering is currently under development. The direct benefits of this component are definitions of and the relations between the most salient planning, management and evaluation processes. These specifications will be useful in precisely defining detailed problems, in monitoring movement toward solutions, and their application to areas beyond compensatory education.

3) Study of Department Processes. This component includes documenting the processes and structures actually operating in compensatory programs in the Department. This study addresses the period 1969-1973 with a view toward appraising the impact of Department policies, guidelines and regulations, procedures and technical assistance on local district planning, management and evaluation practices and project performance. Interviews and analysis of in-house documents, currently being conducted, are yielding the information necessary to complete this task. Direct benefits of this component will be a detailed longitudinal (i.e. multi-year) description of the Department's activities concerning compensatory education.

Question #3: To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?

4) Study of Local District Project Documentation. This component is intended to systematically measure various aspects of local district planning, management, and evaluation practices of compensatory projects on the basis of the applications and the evaluations that districts submit to the Department. This effort covers the period 1969-73, as does the study of Department processes. Data collection and reduction are completed and some of the analyses of these data have been completed. The methods used and the preliminary analyses completed to date are described below. This component is providing two direct benefits. The first is an empirical basis for more detailed understanding of local district planning, management and evaluation processes, the beginning of which is reported below. The second benefit is the development and testing of a series of instruments and procedures for measuring various qualitative aspects of local district planning, management and evaluation processes that could be incorporated into ongoing Department project approval, monitoring, and evaluation procedures, [1] thereby making them more objective and systematic.

5) Department/Local Level Modeling. This component will produce a preliminary descriptive/predictive model of the combined Department/Local Level (within the framework provided by the functional specifications) from the results of the Department description and the local level document study described above. This preliminary model will be "field verified" and refined through interviews in selected local school districts. The model will provide an assessment of the impact of Department actions on local performance, definition of the specific problems in the combined system, and means by which systematic, relatively precise, predictions of probable outcomes of various policy and program options can be made.

1/ Or those of any other State or Federal Agency

Question #3: To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?

6) Trade-off Analyses. This final component will include alternative solutions to the problems identified in the previous section, and a comparison of these alternatives with respect to their feasibility and cost-effectiveness. This appraisal will provide the basis for recommending both policies and programs for the Department as well as workable strategies for local school districts and the Department to improve the planning, management, and evaluation of compensatory education at their respective levels.

Within this context, the information to follow is an interim report of the findings of the fourth component. The next section describes the design, methods and procedures used in the study of local level project documentation.

b. Procedures Used in the Study

The objective of this part of the Project (i.e. the Study of Local District Project Documentation) was to examine the effects of different types of local school districts, programs, and time (independent variables)[2/] on various aspects of local district planning, management and evaluation processes used in compensatory education projects (dependent variables). Three district types were considered: New York City Community School Districts (CSD's), Large Upstate local school districts (i.e. districts outside New York City with more than 15,000 students), and Moderate Upstate local school districts (i.e. districts outside New York City having fewer than 15,000 students but receiving Urban Ed grants greater than $40,000 in 1971-72 [3/]).

2/ As discussed later, additional independent variables will be used in further analyses of these data. The effects of only these independent variables are discussed in this interim report.

3/ The $40,000 grant size cut-off eliminated all districts with enrollments less than 5,000 (small upstate) from the population.

Question #3: To what extent are the best instructional planning, management and evaluation practices being used at different levels of the educational system?

Three programs were included in the study, the State Urban Education program, ESEA Title I, and ESEA Title III. Urban Education and Title I both provide funds on the basis of pre-established formulas under approval of the Department, but differ in the source of the funds (State and Federal, respectively), in the operative guidelines, and in certain aspects of Department processing. Title III is a grant program in which districts compete for funds on the basis of their applications. Comparison of the two formula grants permits appraisal of the effects of funding source, different guidelines, and Department administrative processes. Title III was used as an independent variable because its presence might indicate more aggressive local school district management (i.e., success in competing for funds).

The time dimension for measurements taken from applications extends through the four school years, 1969-70 to 1972-73; but those from evaluation documents cover only three years, 1969-70 to 1971-72, because evaluation documents for 1972-73 were not available when this activity was undertaken in the summer of 1973.

Twenty-five districts containing a sample of 100 one-year E.S.E.A. Title I projects, 100 one-year Urban Education projects and 15 three-year E.S.E.A. Title III projects were selected for analysis in the study. Districts were selected based on the following criteria: (1) they operated both Urban Education and Title I programs and (2) their 1971-72 Urban Education grants were greater than $40,000. The resulting group of 49 districts was then divided into those having Title III programs and those not having Title III. Fifteen of seventeen districts having Title III projects were selected and ten of the remaining 32 districts were selected to give five districts in each of the following categories: (1) New York City Community School Districts (NYC/CSD) having Title III projects; (2) NYC/CSD's not having Title III projects; (3) moderate-sized upstate districts having Title III; (4) moderate-sized upstate districts not having Title III; and (5) large upstate districts (all of which also had Title III projects). This 25-district sample was used in the study.

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

In selecting individual projects for each of these twenty-five districts, the following criteria were applied: the project was (1) continuous in terms of subject matter over the 1969-73 period, (2) funded at a level greater than $40,000, (3) cognitively oriented (generally reading or math), and (4) aimed at elementary grade students. If one of the criteria could not be met specific decision rules were followed for selection.

An instrument was developed specifically for this study to measure the nature and quality of local school district instructional planning, management and evaluation processes through application and evaluation documents. It was pilot tested in March, 1973, reviewed, discussed and revised.

The resultant instrument was composed of five groups of scales measuring planning, management and evaluation performance and one group of scales measuring responsiveness to program guidelines.

Eleven persons were hired for the task of rating the local school district documents. These project raters were given intensive eight-day training in the use of the instrument which involved a complete review of the instrument and scoring procedures and a two-day sample scoring exercise.

The study instrument was applied to 215 project applications and 159 project evaluations from three programs: ESEA Title I, Title III, and Urban Education.

As indicated above, the instrument was composed of five groups of scales, three of which were included in this analysis: Application Quality (21 scales); Evaluation Quality (15 scales); and Responsiveness to Guidelines (58 scales Title I, 22 and 13 scales for 1969-7 and 1971-73 Urban Education respectively); which were repeated a number of times for each district.

Question #3: To what extent are the best instructional planning management, and evaluation practices being used at different levels of the educational system?

To facilitate analysis and interpretation related scales were combined to form "indicators" of more complex behaviors. Preliminary analysis included the following indicators constructed by grouping individual scales that relate to the behaviors under study:

1) Application Quality (21 scales)

2) Title I Responsiveness (58 scales)

3) Urban Education Responsiveness, 1969-71 (22 scales)

4) Urban Education Responsiveness, 1971-73 (13 scales)

5) Evaluation Quality (15 scales)

Because time did not permit a second pilot test of the instrument, a reliability substudy was conducted. Five scorers each scored the same three E.S.E.A. Title I projects and the Responsiveness sections of the same four Urban Education projects using standard scoring procedures. The major reliability issue was the extent to which different scorers used the scales in the same ways. To assess this, inter-scorer reliability coefficients[5/] were computed for each study and are shown in Table 46. The coefficients are high enough to permit the conclusions that the indicators are reliable for the present purposes.

5/ The Pearson product-moment correlation coefficient (r) was calculated for each pair of scorers for each indicator yielding 10 correlation coefficients each. These 10 coefficients were then "averaged" by transforming each of them to Fisher's Z coefficient, summing these scores, dividing by 10^{19} and reconverting the averaged coefficient Z to r.

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

Table 46

Reliability Coefficients of Indicators

Indicator	Reliability Coefficient
1) Application Quality	.81
2) Title I Responsiveness	.55
3) Urban Education Responsiveness, 1969-71	.73
4) Urban Education Responsiveness, 1971-73	.47
5) Evaluation Quality	.88

The Application Quality scales deal with how well a district describes its proposed project in the application document submitted to the Department. The application quality indicator consists of 21 scales rating documents based on the following major groupings of criteria:

(1) Needs - the precision with which the application describes district-wide and target group needs of educationally disadvantaged students.

(2) Objectives - the measurability of objectives and the degree to which objectives are stated in student behavioral outcome terms.

(3) Activities - the detail with which activities are specified.

(4) Relationships - the degree to which objectives address needs and the degree to which activities address objectives.

(5) Planning - the detail of description of the process by which the project was planned including the specificity of the consideration of alternatives.

(6) Participants - the degree to which criteria used to select participants (professional staff, nonprofessional staff and students) are project relevant.

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

(7) Evaluation - a) the extent to which the planned evaluation addresses the stated project objectives; b) the extent to which issues of instrument validity and reliability are discussed; c) the precision of the evaluation design description; d) the relevance of comparisons proposed in the evaluation design.

The Title I Responsiveness indicator consists of 58 scales which are designed to measure a district's responsiveness to official program guidelines. The scales vary from evidence of non-compliance to evidence of complete compliance with a selected sample of Title I guidelines* (evidence of partial compliance is the midpoint on the scale). Title I program guidelines generally consist of criteria by which applications are considered for approval. The 58 scales included in the Title I Responsiveness section, then, are based on selected groups of approval criteria. These groups include:

(1) Selection of attendance areas;

(2) Assessment of needs of disadvantaged children;

(3) Planning and coordination of Title I program with other programs and agencies;

(4) Program design considerations such as concentration of funds on high priority needs, statement of program objectives and participation of private school students;

(5) Program implementation considerations such as selection and use of staff, staff training, parent involvement, use of funds for instruction and facilities and dissemination of information;

(6) Provisions for evaluating the program;

(7) Comparability of services for disadvantaged children; and

(8) New York State Title I priorities.

---

* Guidelines used in the Title I Responsiveness section of the instrument are principally based on Title I Program Guide #44 (INST. L203.1).

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

The Urban Education Responsiveness Indicator for 1969-71 consists of 22 scales which attempt to measure a district's responsiveness to official Urban Education Program guidelines in operation from 1969 to 1971. The Urban Education guidelines for 1969-71 generally prescribe the types of information which should be included in an application. The 22 scales included in the Urban Education Responsiveness section for 1969-71, then, are based on the types of information prescribed in the guidelines such as:

(1) Statement of objectives;

(2) Description of target group(s);

(3) Description of the geographic area to be served;

(4) Description of the proposed project including number and type of pupils, roles of personnel, number and type of staff, costs, etc.;

(5) Description of means for attaining coordination and achieving community participation;

(6) Description of evaluation and dissemination procedures.

The Urban Education Responsiveness Indicator for 1971-73 consists of 13 scales which attempt to do the same thing as the 1969-71 scales using a revised set of guidelines. These revised guidelines also tend to prescribe, in certain instances, the types of information that should be included in an application (e.g. description of target groups). However, the revised guidelines include certain program limitation guidelines such as:

(1) Limiting evaluation expenditures to 5%;

(2) Limiting program administration expenditures to 10%;

(3) Establishing minimum cost per participant at approximately $150;

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

(4) Establishing the ratio of teacher aides to participants at less than 1:20;

(5) Determining that Urban Education funds should not be used to supplant federal, local or other state funds previously supporting a program.

Evaluation quality scales deal with how well a district describes its proposed project in the evaluation document submitted to the Department. The evaluation quality indicator consists of 15 scales rating documents on the following major groupings of criteria:

(1) Objective - the measurability of objectives, the degree to which objectives are stated in student behavioral outcome terms and the extent to which the evaluation addresses the project objectives.

(2) Evaluation Design - the precision with which the experimental design used in the evaluation is described and the appropriateness of comparisons proposed in the design as presented in the evaluation document.

(3) Activities - the detail of the description of project activities that were carried out.

(4) Participants - the detail given on the characteristics of professional staff, paraprofessional staff and student participants in the project.

(5) Data Analysis - the extent to which instrument validity and reliability is described, the extent to which the evaluation document presents evidence of bias in the sampling of students to be tested, and the degree to which data are summarized and subsetted.

(6) Conclusions - the extent to which conclusions and recommendations are directly supported by data and are relevant for modifying the project in the future.

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

c. Preliminary Results of Study

The findings reported below related only to the application and evaluation documents, and not necessarily to the actual program operations in the field.

The Application Quality indicator measures the extent to which local school district applications approach contemporary management and planning standards. High quality applications, those scoring near 1 on the indicator, would closely approximate the best that might be expected; low scores, near 0, would indicate that the applications are far from planning theory norms. The applications being scored are the products of local school district planning efforts, and thus the Application Quality indicator may be interpreted as a "proxy" measure of planning quality.

The mean Application Quality scores over all districts and years was as follows: Title I, .51; Urban Education, .48; and Title III, .51. A comparison of the three programs for all three district types (15 local school districts), found no significant differences in Application Quality scores among programs or district types (see Table 47).

Table 47

Mean Application Quality Scores
by District and Program Type

District Type	Title I	Urban Education	Title III
Moderate Upstate	.598	.458	.508
Large Upstate	.480	.538	.474
NYC/CSD's	.490	.492	.558

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

A statistically significant trend of improving scores for both Title I and Urban Education projects over all local school district types from 1969-70 to 1971-72 (see Table 48) was noted, with a leveling off occurring in 1972-73. This trend, however, only represented a gain from a generally poor level of quality in 1969-70 to an overall fair level of quality in 1972-73. The NYC/CSD's showed the most marked improvement over the time period, starting at a lower level (.349) but improving consistently to .565 over the time period. The moderate size upstate districts began at a higher level (.434), improved less rapidly, and actually dropped very slightly in the last year to .518 (see Figure 3). In addition, no significant overall differences were noted between the Urban Education and Title I programs over the time period. However, for moderate upstate districts, Title I applications had higher scores than Urban Education applications, while the opposite was true for large upstate districts. No difference was observed between programs for the NYC/CSD's (see Figure 4).

Table 48

Combined Title I and Urban Education Application Quality Scores Over Time

	1969-70	1970-71	1971-72	1972-73
Combined Title I and Urban Education scores for all district types	.407	.483	.544	.553

Overall, despite improvements in some district types, application quality scores were relatively low over the four-year period under study when compared to contemporary planning standards. Further interpretations of the significant differences noted above will only be made after further analysis of the data.

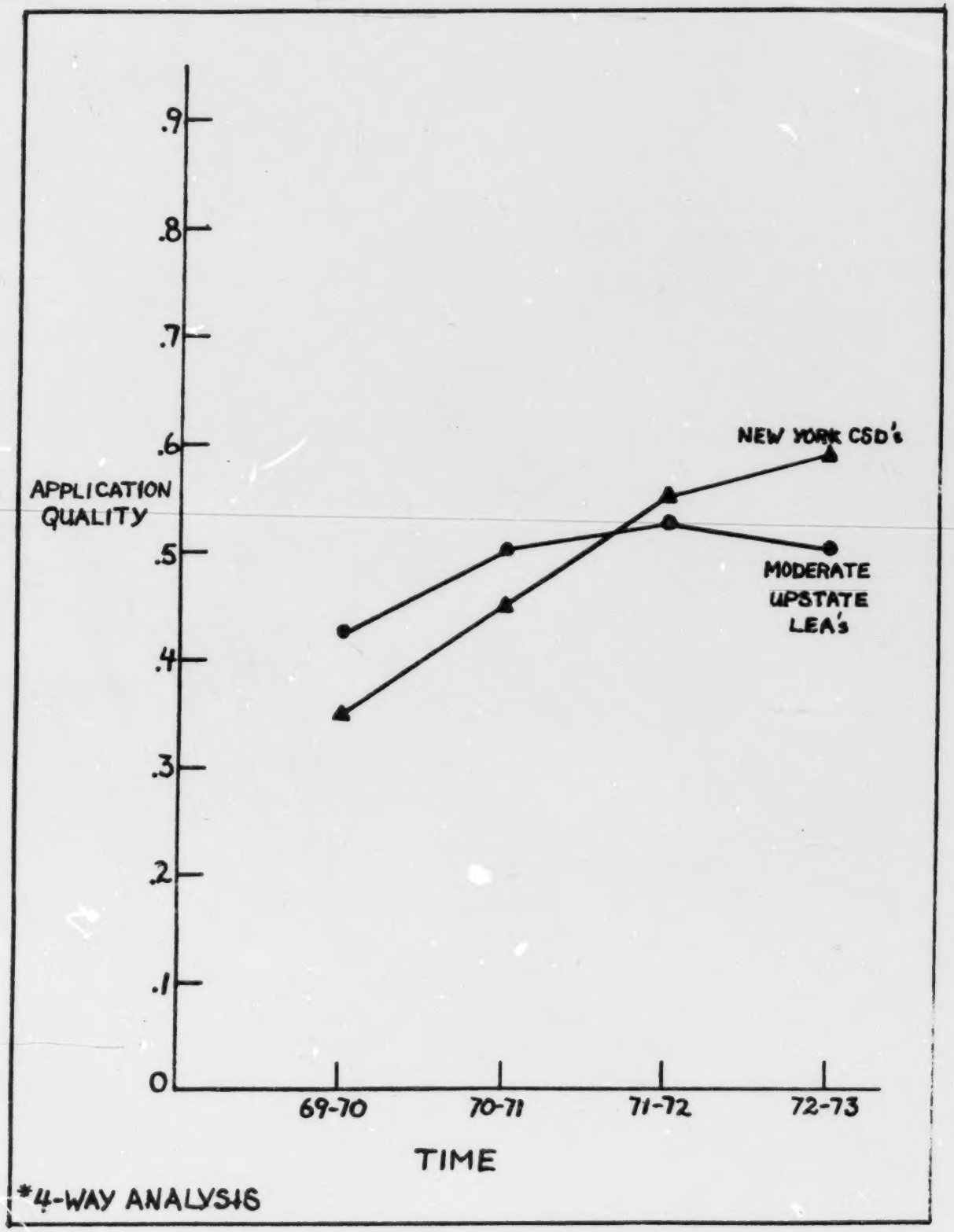


FIG. 3 APPLICATION QUALITY BY DISTRICT TYPE AND TIME*

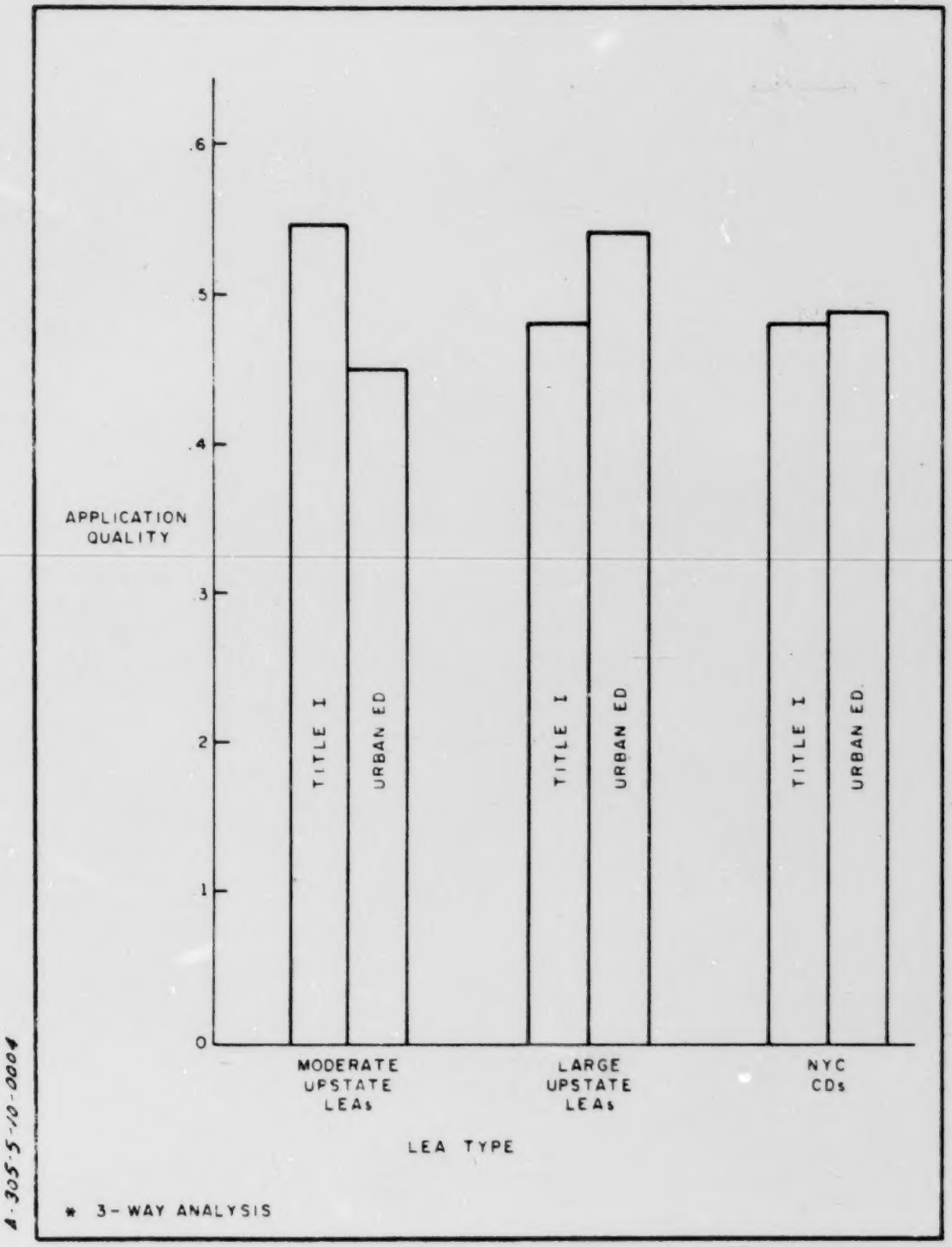


FIG. 4 APPLICATION QUALITY BY DISTRICT TYPE AND PROGRAM*

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

The Responsiveness indicator measures the extent to which project applications reveal conformance to the guidelines and regulations operative in the program and year in which the applications were submitted. The application themselves (not necessarily the operating projects) must describe behavior in conformance with the regulations. Explicit mention of the guideline or descriptions of conforming behavior in the documents is interpreted as evidence that the districts have acknowledged the guideline in planning the project. This interpretation is based on the fact that these documents are the principal source of information by which the Department project approval process verifies local school district conformity with guidelines. The guidelines and regulations differ between the programs; the Urban Education guidelines were revised during the period of this study so separate indicators were used for Title I, 1969-71 Urban Education, and 1971-73 Urban Education projects.

The mean combined responsiveness score for all districts and years was .644, indicating a moderately high degree of overall conformance to guidelines and regulations.

When considered over time the combined responsiveness scores for the New York Community School Districts and moderate size upstate local school districts were about the same in 1969-70 and 1970-71. During the last two years of the study period, however, the New York city district's Responsiveness Scores improved, while the moderate upstate district's responsiveness scores dropped.

When the Title I and Urban Education programs were examined separately, the following mean scores were reported: Title I Responsiveness for all districts and over all years .59; Urban Education Responsiveness for all districts from 1969-71 .68; Urban Education Responsiveness for all districts from 1971-73 .69. The above means represent statistically significant differences in the two programs with Urban Education scores generally higher over the four-year period than Title I scores. Although the scores of the two programs were significantly different

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

statistically, it must be noted that substantial differences exist in the sets of guidelines, and therefore, scales used to measure responsiveness to those guidelines, result in different scores. Title I guidelines tend to emphasize program details and criteria by which applications will be considered for approval. By contrast, the Urban Education guidelines are more oriented to procedural issues (e.g. how to format applications, what information to include in applications, etc.). Consequently, districts which score high in regard to the guidelines of one program may not necessarily score high in relation to the guidelines of the other program. Analysis involving program and district type revealed significant findings consistent with the above hypothesis. For example, moderate upstate districts had about the same level of responsiveness on both Title I and Urban Education over the four-year period. However, the New York CSD's scored lower on Title I than on Urban Education. Additional comparisons between programs must await further analysis.

Overall, responsiveness indicators showed a moderately high degree of conformity between applications submitted by local school districts and the operative program guidelines in the two programs under study.

The Evaluation Quality indicator measures the extent to which the evaluation documents submitted by districts to the Department approach contemporary project evaluation standards. High quality evaluations (scores approaching 1.0) would clearly approximate the best that might be expected; low scores (near zero) would indicate very poor evaluation performance. Since the evaluations scored are either the products of districts' own efforts or the products of contractors working under the supervision of local officials, the Evaluation Quality indicator may be interpreted as a "proxy" measure of the quality of districts' evaluation processes or capabilities.

The mean Evaluation Quality indicators for all local districts and time periods were .44 for Title I and .48 for Urban Education, indicating a rather poor quality of evaluation documents. A significant upward trend in combined

Question #3: To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

Evaluation Quality scores from .416 to .492 was found over the three years for which evaluation documents were scored. However, this overall improvement can be principally attributed to the increase in New York CSD's which had an extremely low mean score of .32 in 1969-70 and moved to a fair level of .51 in 1970-71. Both the large and moderate upstate districts remained relatively the same, the means ranging closely around .45 for all three years. Table 49 and Figure 5 represent the combined Evaluation Quality Scores for the three district types over the study period.

Table 49

Combined Title I and Urban Education Evaluation Quality Scores by District Type Over Time

	1969-70	1970-71	1971-72
Moderate Upstate	.456	.452	.467
Large Upstate	.470	.432	.474
NYC/CSD's	.321	.514	.534

In summary, Evaluation Quality was found to be fairly stable at a relatively low level over the three-year period. Title I and Urban Education projects scored about the same on this variable. Significant improvement over the three-year period was noted for New York CSD's.

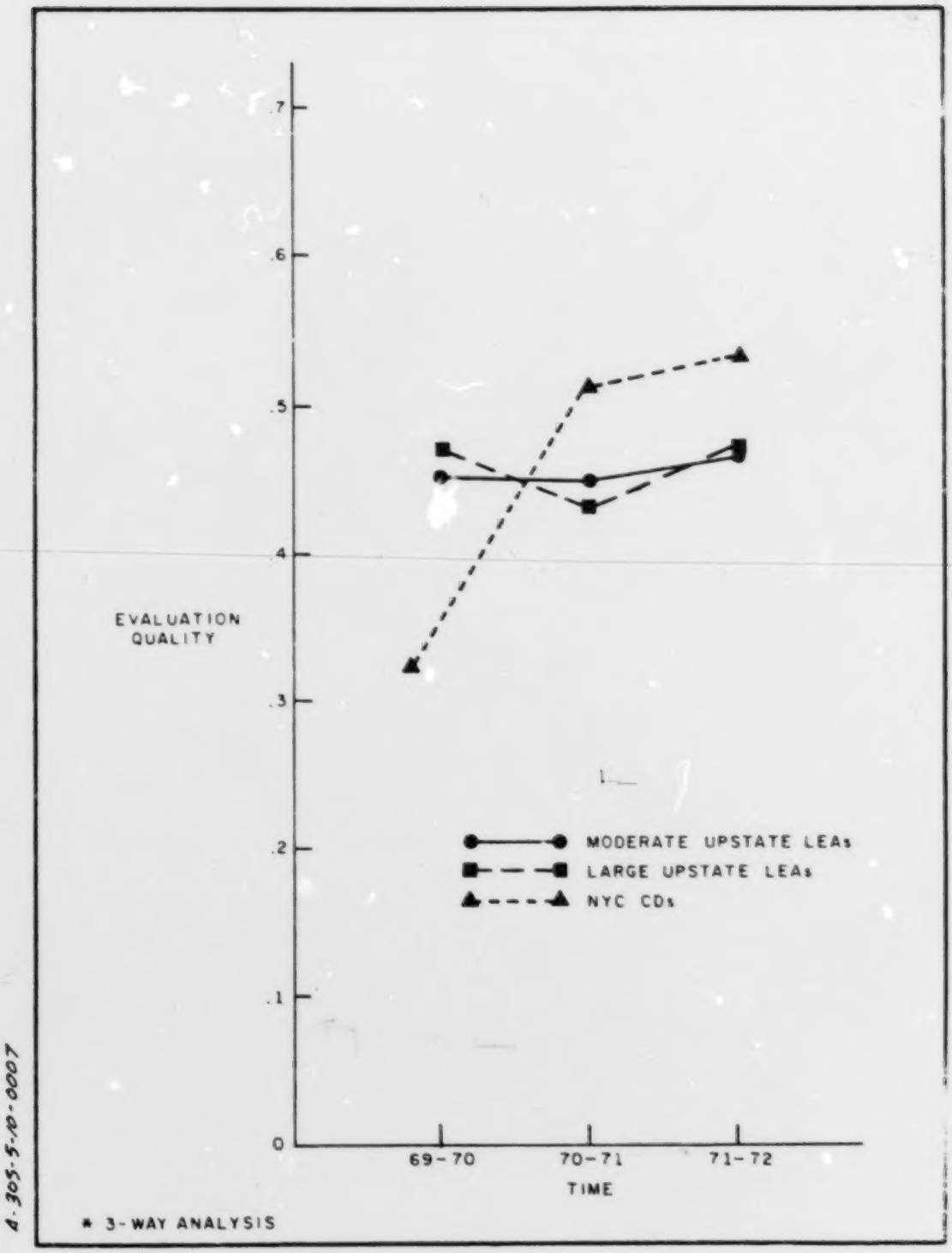


FIG. 5 EVALUATION QUALITY BY LEA TYPE AND TIME *

Question #3. To what extent are the best instructional planning, management, and evaluation practices being used at different levels of the educational system?

d. Limitations of these Data

As indicated above only preliminary analysis of these data has been performed. Further analysis is currently underway. All findings will be followed up and substantiated through field visits to local districts. An intensive, structured interview is planned to determine: (1) the planning and evaluation processes as they exist at the local level and (2) the correspondence between those processes and the processes reported in the Title I and Urban Education application and evaluation documents submitted to the State Education Department.

e. Summary of Findings

Responsiveness scores were relatively high in both the Urban Education and Title I programs, in all local school district types and in all study years. General upward trends in scores over the four-year period were noted for Application Quality scores for all district types in the two programs. Similar improvement was found over the study time period in the Evaluation Quality scores for New York City community districts in both programs. Evaluation Quality scores for large and moderate upstate districts for both programs remained relatively stable over the three-year period. Despite existence of these improvements in scores, both Application and Evaluation Quality scores were still found to be quite low when compared to contemporary planning and evaluation standards.

These findings represent only a preliminary analysis of the data. Intensive field interviews will be conducted in order to describe actual local school district operations more precisely. Further analysis and comparisons of data obtained from both the scoring of documents and the interviews will be performed prior to making recommendations in fall 1974.

APPENDIX A

DESCRIPTION OF ACTIVITIES THAT MAKE I. E. S. AN OPERATIONAL SYSTEM

A. Spring 1974 - Workshops

1) Formulate or revise program (e.g. reading) and subprogram (e.g. grade 1) objectives

2) Divide subprograms into modules

3) Form banks of instructional activities (employing specified or new, teacher-generated materials) linked to objectives

Activities

Three types

(1) "First-try" educational processes

(2) "Compensatory" processes

(3) "Enrichment" processes

4) Generate mastery test items to measure the extent to which instructional objectives have been met by students

5) Determine the particular type and extent of individualized instruction to be employed

6) Determine the organizational means of implementing instruction, e.g. self-contained classrooms, team teaching, etc.

7) Set mastery standards, i.e., what proportion of test items associated with each objective must be correctly answered to establish mastery of that objective.

8) Decide on roles and missions of all personnel connected with the program

For example:

--From whom can a teacher seek help when it is needed?

--Who is responsible for takin action if students do not meet district standards?

B. Summer 1974 - Preparation for Implementation

1) Before the school year begins, mastery and retention tests for all modules must be prepared and inventoried

2) Machine-readable mastery test answer sheets or forms must be produced

3) Program manuals describing modules, objectives, instructional activities (employing designated materials), and performance criteria for each objective must be made up and distributed

C. Beginning of the School Year 1974-75

1) Students are assigned to teachers

2) Elements of pupil, personnel and program files necessary to implement ISS are set up

3) Individually addressed placement tests (with student's name, I.D. number, school and teacher's name pre-entered) are given to students to enable teachers to determine which objectives will be taught to different students or groups of students

4) Test request forms listing all students assigned to each teacher are prepared and delivered to each teacher together with all other ISS materials

D. During the School Year 1974-75

1) Instructional activities and resources are selected from the banks by teacher (with or without student participation) and delivered to individual students or to groups of students by teachers

2) As each student or group of students nears the end of a module, the teacher files a test request form, indicating the required test number and checking off the students to be tested

3) In a few days, individually addressed mastery tests (with answer sheets) are delivered to the teacher

4) The teacher administers the tests and returns the answer sheets for scoring

5) Within twenty-four hours, test scores (by objective) are reported to the teacher and to the students

6) The teacher modifies instruction (e.g., selects a set of activities) based on a diagnosis of test score results

7) After some time, retention tests for each module are requested, administered, scored, and reported

8) All records of test scores and instructional activities administered are kept in computerized files for analysis and evaluation of module and subprogram effectiveness.

9) Monthly summary and exception reports of student performance are sent to administrators. Other reports can be generated on request

E. <u>After the School Year is Over</u>

1) An evaluation of overall district performance by program is executed, and a report having a form and content specified by the district is generated

2) Analyses of pupil and program files are made to identify ineffective activities as a function of learner characteristics

3) Analyses of the differential effectiveness of activities by learner characteristics are executed and a decision is made to enter the results in descriptive language into the district's ISS program books (e.g., for a particular activity, the entry might be "works particularly well for students who speak non-standard English")

4) Analyses to support budget-making and resource allocation decisions are executed as required